On the Role of Shared Randomness in Simultaneous Communication*

Mohammad Bavarian† Dmitry Gavinsky ‡ Tsuyoshi Ito §

Abstract

Two parties wish to carry out certain distributed computational tasks, and they are given access to a *source of correlated random bits*. It allows the parties to act in a correlated manner, which can be quite useful. But what happens if the shared randomness is not perfect?

In this work, we initiate the study of the power of different sources of shared randomness in communication complexity. This is done in the setting of *simultaneous message passing (SMP) model of communication complexity*, which is one of the most suitable models for studying the resource of shared randomness. Toward characterising the power of various sources of shared randomness, we introduce a measure for the "quality" of a source – we call it *collision complexity*. Our results show that the collision complexity tightly characterises the power of a (shared) randomness resource in the SMP model.

Of independent interest is our demonstration that even the "weakest" sources of shared randomness can in some cases increase the power of SMP substantially: the *equality* function can be solved very efficiently with virtually any nontrivial shared randomness.

1 Introduction

One of the central themes of complexity theory is the study various resources of computation and their respective power. In this direction many different models of computation have been defined, and various resources have been considered – e.g., time, space, randomness, access to powerful provers, etc. Among the strongest models where the researchers have been able to prove strong lower bounds is the model of *communication complexity*.

An important resource in communication complexity and, more generally, in virtually all distributed computational settings, is the access to a source of *correlated random bits* or *shared randomness (SR)*: It allows the parties to act in a correlated way in order to perform better and more efficiently in their tasks.

The primary goal of this work is a quantitative characterisation of the utility of different sources of shared randomness to parties interested in solving communication problems. To this end, let us define sources of shared randomness more formally.

*A preliminary version of this article appeared in the proceedings of ICALP 2014, pages 150-162.

†Massachusetts Institute of Technology, Cambridge, MA, U.S.A. Partially supported by NSF STC Award 0939370. Part of this work was done while at the NEC Laboratories America, Princeton, NJ, U.S.A.

‡Institute of Mathematics, Czech Academy of Sciences, Praha, Czech Republic. Partially funded by the grant P202/12/G061 of GA ČR and by RVO: 67985840. Part of this work was done while at the NEC Laboratories America, Princeton, NJ, U.S.A.

§Part of this work was done while at the NEC Laboratories America, Princeton, NJ, U.S.A.

Definition 1.1 (Source of SR for two parties)**.** A source of shared randomness (also referred to as a bipartite distribution) is specified by a distribution ρ on a domain $U \times V$, where U and V correspond to the parts of shared randomness visible to Alice and to Bob, respectively. To solve their communication task, the parties are allowed to take as many independent samples as they need from the source ρ.

To understand the motivation behind the above definition better, it would be useful to consider a few examples of sources of shared randomness. Let $U = V = \{0, 1\}$. The following two sources on $U \times V$ are the well-known settings of *private* and *perfect* forms of shared randomness, respectively:

$$\rho_{priv} := \mathcal{U}_{\{00,01,10,11\}}, \qquad \rho_{perf} := \mathcal{U}_{\{00,11\}},$$

where $\mathcal{U}_{\{...\}}$ is defined by:

Notation 1.2. For a finite set Q, we denote the uniform distribution over it by $\mathcal{U}_Q$.

The above two examples (ρ_{perf} and ρ_{priv}) are valuable as they indicate that the more familiar settings of communication complexity are given as special cases of our model by setting $\rho = \rho_{priv}$ or $\rho = \rho_{perf}$. However, more interesting questions in this work are related to the intermediate cases between the "extremes" of ρ_{perf} and ρ_{priv}.

1.1. Shared randomness in SMP communication: definitions and more examples

Communication complexity studies how much communication is needed to compute a given function on distributed inputs. The model we work with is the two-party *simultaneous message passing (SMP)* model, in which the *players*, *Alice* and *Bob*, each send a message to the *referee* [19, 2]. The SMP model augmented with a shared randomness source ρ is defined as follows: Let ρ a source of shared randomness on $U \times V$ and $(u_1, v_1), (u_2, v_2), \ldots, (u_m, v_m)$ be independent samples from ρ. In the SMP model with shared distribution ρ, Alice receives input x and her part of shared randomness $R_a = (u_1, \ldots, u_\ell) \in U^\ell$, and Bob receives y and his part of shared randomness $R_b = (v_1, \ldots, v_\ell) \in V^\ell$. Alice and Bob use their parts of input and shared randomness to compute their messages, $M_a(x, R_a)$ and $M_b(y, R_b)$, to the referee who upon receiving these messages outputs an answer. A *communication protocol* determines the value of m and the actions of all the participants in the protocol.

A communication protocol is said to solve a communication problem with error probability δ if it guarantees correct answer with probability at least $1-\delta$ for every allowed input. The communication cost of a protocol is the maximum possible total number of bits sent by the players to the referee. Given a communication problem f, we denote the minimum cost of a protocol that solves f (with error probability $< \frac{1}{3}$) in the SMP model with SR source ρ by $SMP_\rho(f)$.

Our goal. In this work we consider the case when ρ is not a perfect source of shared randomness and investigate to what extent Alice and Bob can still use ρ to reduce the communication cost (as compared to the case of private randomness). Moreover, we are interested in understanding *what properties of the source ρ determine its utility as a resource in the SMP model.* Consider the following slightly more complicated distributions.

Example 1.3. Let $U = V = \{0, 1\}$.

- **"Shared disjointness"**: The shared distribution is $\rho_{disj} := \mathcal{U}_{\{00,01,10\}}$.

- **Symmetric Noisy Bits**: The shared distribution is $\rho_{BSC_p} = (1-2p)\rho_{perf} + 2p\rho_{priv}$, i.e. $\rho_{BSC_p}(0,0) = \rho_{BSC_p}(1,1) = \frac{1-p}{2}$ and $\rho_{BSC_p}(0,1) = \rho_{BSC_p}(1,0) = \frac{p}{2}$.

The subtlety of the problem of determining the power of a general source of shared randomness ρ becomes more clear when one notices that even in the simple case of ρ_{disj} vs. ρ_{BSC_p} it is a priori unclear which source is stronger than the other (say, for $p = 0.8$). Even more, it is not obvious that it should possible to make a useful comparison between two different sources of shared randomness ρ_1 and ρ_2 which, like the above case of ρ_{disj} vs. ρ_{BSC_p}, may have quite different forms.[1] Maybe surprisingly, it turns out that comparison between sources of shared randomness of very different form is possible (see Theorem 1.8). To achieve this, we introduce the notion of *collision complexity* of a source (see Section 5), which associates to a source ρ a real valued function $\mathrm{col}_\rho : \mathbb{N} \to \mathbb{R}^+$, whose rate of growth captures the utility of ρ as a source of shared randomness.

Before we continue, let us exclude the following case:

Definition 1.4. A source of shared randomness is *degenerate* if it is deterministic from the point of view of at least one of the players.

In other words, a source of shared randomness is non-degenerate if the players can use it to *locally emulate private randomness (ρ_{priv})*. Degenerate sources are not quite that interesting for the purposes of this work as they are effectively no more powerful than the deterministic ones, and therefore throughout the paper all sources of shared randomness are assumed to be *non-degenerate*.

1.2. Our results

The first result is an example of an SMP protocol showing that even the weakest sources of shared randomness can be sometimes quite powerful. More specifically, we have the following theorem for the equality function $EQ_n \colon \{0,1\}^n \times \{0,1\}^n \to \{0,1\}$, defined to be 1 if $x = y$, and 0 otherwise.

Theorem 1.5. *Let ρ be a probability distribution on $U \times V$. If ρ is not a product distribution, then it holds that $SMP_\rho(EQ_n) = O(1)$, where the implied constant in $O(\cdot)$ depends only on ρ (and not on n).*

The above result may look somewhat surprising, as we have $SMP_{priv}(EQ_n) = \Theta(\sqrt{n})$ [2, 19], which means that even the presence of slightest correlation in the shared distribution ρ increases the power of SMP significantly in the asymptotic sense, decreasing the communication requirements from $\Omega(\sqrt{n})$ to $O(1)$.

The above theorem gives rise the following question:

Question 1.6. Is it possible that what happened for EQ_n is generic? More precisely, is the effect of having an imperfect source of shared randomness ρ instead of ρ_{perf} limited to at most a multiplicative constant blow-up in the communication, i.e.

$$SMP_\rho(f) = O_\rho(SMP_{perf}(f))$$

for any fixed non-product ρ?

Note that since the constant in $O(\cdot)$ in the above may depend on ρ (but not on the instance size n), the above question is not as naive as it may at first appear. For instance, if one trusts the

[1] Compare this to the case ρ_{BSC_p} vs. ρ_{BSC_q} where evidently when $|p| < |q|$ the first source is more powerful than the second as the players can always use private randomness to locally simulate ρ_{BSC_p} using ρ_{BSC_q}.

(incorrect) intuition that "usefulness" of a source of shared randomness $(A, B) \sim \rho$ is determined by the mutual information between the two sides, $I(A; B)$, then it may appear reasonable to believe in the affirmative answer to Question 1.6.

Nevertheless, the answer to Question 1.6 is *negative*, and the actual trade-off between $SMP_\rho(f)$ and $SMP_{perf}(f)$ is more subtle, and, as shown in Theorem 1.8, is captured by the collision complexity of ρ. Hence, besides the technical significance of Theorem 1.8, we believe that the fact that collision complexity (as opposed to entropic quantities like mutual information, etc.) is *the appropriate measure for quantifying the power of a shared random source* is perhaps a noteworthy conceptual message of the result.

Before stating our main theorem, let us give an (informal) definition of the collision complexity and observe some of its basic properties.

Definition (Collision complexity, informal)**.** For $n \in \mathbb{N}$, we say that a two-player protocol *produces a collision* if each player output a subsets of $[n]$, respectively denoted by A and B, such that $\Pr[i \in A \cap B] \geq 1/n$ for each $i \in [n]$.[2] The complexity of a protocol is the maximum possible $\frac{|A|+|B|}{2}$. The collision complexity of a source ρ, denoted by $\mathrm{col}_\rho(n)$, is the minimum complexity of a protocol that uses (arbitrary number of) samples from ρ and produces a collision.

Conceptually, collision complexity quantifies the cost of simulating perfect shared randomness by two parties who only have ρ, and therefore it can be viewed as *a measure of quality* of ρ. For example, it is clear from the above definition that the players sharing ρ_{perf} can always use their perfect source of shared randomness to agree on an element $i \in [n]$ and both output $A = B = \{i\}$ which shows that ρ_{perf} has collision complexity 1, which is optimal. The following basic properties of collision complexity also follow easily from the definition.

Fact 1.7 (Properties of collision complexity)**.** The collision complexity $\mathrm{col}_\rho : \mathbb{N} \to \mathbb{R}^+$ satisfies the following:

i. **Basic General Bounds**: $1 \leq \mathrm{col}_\rho(n) \leq \sqrt{n}$.

ii. **Private and Public Randomness**: $\mathrm{col}_{\rho_{perf}} = 1$ and $\mathrm{col}_{\rho_{priv}}(n) = \sqrt{n}$.

iii. **Tensor Product Stability**: $\mathrm{col}_{\rho^{\otimes t}}(n) = \mathrm{col}_\rho(n)$.

iv. **Decrease of Collision Complexity after Tensoring**: $\mathrm{col}_{\rho\otimes\sigma}(n) \leq \min\{\mathrm{col}_\rho(n), \mathrm{col}_\sigma(n)\}$.

Combining the property (i) and (ii) from the above, we see that the minimal value of collision complexity is achieved by ρ_{perf} and its maximum value is achieved by ρ_{priv}. The ***main intuition*** here is that the higher the "quality" of the correlation provided by ρ is, the slower we expect the growth of $\mathrm{col}_\rho(n)$ to be.

We can now state our main theorem which uses collision complexity to compare SMP_ρ with SMP_{perf} (where SMP_{perf} is a shorthand for $SMP_{\rho_{perf}}$).

[2] An alternate point of view on this criteria is that the parties want to ensure $\mathbf{E}[|A \cap B|] \geq 1$, and furthermore, they want the intersecting sets $A \cap B$ to be distributed more or less evenly across $[n]$. Of course, without the latter condition achieving $\mathbf{E}[|A \cap B|] \geq 1$ would have been easy, since the players could otherwise have always outputted a fixed $i \in [n]$; however, the evenness condition along with the desire to minimise $|A| + |B|$ makes this a non-trivial task.

Theorem 1.8. *For any communication problem f and non-degenerate source of shared randomness ρ, we have*

$$SMP_\rho(f) = \widetilde{O}(SMP_{perf}(f) \cdot \mathrm{col}_\rho(n)\,). \tag{1}$$

Moreover, there exists a family of partial function $\{g_n\}$ such that

$$SMP_\rho(g_n) = \widetilde{\Theta}(\mathrm{col}_\rho(n)\,). \tag{2}$$

The first part of Theorem 1.8 states that any source of correlation can "replace" the perfect source $\rho_{perf} = \mathcal{U}\{00, 11\}$ at the multiplicative cost of $\mathrm{col}_\rho(n)$. The second part of the theorem guarantees the existence of a family of partial functions with $SMP_\rho(g_n)$ exhibiting the behaviour expected from the first part of the theorem – which proves tightness of the result, up to poly-logarithmic factors.

The particular family of partial functions in the second part of Theorem 1.8 is defined as follows.

Definition 1.9 (Gap-inner-product)**.** Let $n, m \in \mathbb{N}$ and $\forall i \in [n] : x_i, y_i \in \mathbb{F}_2^m$. Define the *gap-inner-product function* $GAPIP_{n,m}$ as

$$GAPIP_{n,m}((x_1, \ldots, x_n), (y_1, \ldots, y_n)) = \begin{cases} 0, & |\{i \in [n] \colon x_i \cdot y_i = 0\}| \geq 2n/3, \\ 1, & |\{i \in [n] \colon x_i \cdot y_i = 1\}| \geq 2n/3, \\ \perp, & \text{otherwise.} \end{cases}$$

We set $g_n = GAPIP_n$ where $GAPIP_n$ is shorthand for $GAPIP_{n,8\log n}$, assuming that n is a power of two in this context.

Note that the second part of Theorem 1.8 itself consists of both an upper bound and a lower bounds on $SMP_\rho(GAPIP_n)$ (the main technical challenge being proving the lower bound).

Our last set of result (presented in the appendix) describes the relation between the rate of growth of $\mathrm{col}_\rho(n)$ and the hypercontractivity properties of ρ, using that to give bounds on the collision complexity of ρ_{disj}.

2 Background and related work

Simultaneous message passing model. As we alluded to before, the SMP model is particularly suitable for studying shared randomness. This is due to the fact that the power of this model depends crucially on availability of a common random source. It is easy to see that shared randomness allows the players to use *mixed strategies*.

The classical example of *the equality function* on n-bit strings demonstrates a problem that requires $\Theta(\sqrt{n})$ bits of communication in the SMP model with private randomness only, but can be solved by a $O(1)$-bit protocol in the SMP model with (perfect) shared randomness. Moreover, we will see in Section 4 that *any non-trivial form of shared randomness* is sufficient for solving the equality function by a $O(1)$-bit protocol. On the other hand, in virtually all other existing two-party models, at most additive logarithmic "saving" in terms of communication complexity can result from the availability of shared randomness (as opposed to private randomness only).

In this work we generally view poly-logarithmic multiplicative factors as insignificant in the context of communication complexity, and therefore for our needs SMP model is the most suitable model to consider. [3]

[3] On the other hand, in a parallel work, Canonne et al. [4] study the effects of shared randomness in the one- and two-way settings in the "sub-logarithmic regime" and obtain rather interesting and unexpected results.

Collision complexity and measures of quality of correlation. Intuitively and informally, our notion of collision complexity is *a measure of quality of shared randomness* ρas it satisfies the tensor product stability property (Fact 1.7). A number of other "measures of quality" of correlations have been extensively studied. Perhaps the most well-known among those is the *maximum correlation* and *hypercontractivity*. The latter which is of particular interest due to both its utility in a multitude of applications and its striking mathematical elegance.

The literature on measures of quality of shared randomness (and in particular hypercontractivity) is vast and we will not try to give a comprehensive survey here; instead, we refer the reader to recent works [1, 7, 13] and references therein. We just briefly note one particular interesting line of work on a related problem called *non-interactive correlation distillation (NICD)* [15, 17, 22]. In the two-party NICD, Alice and Bob have access to an unlimited number of independent copies of ρ (just as in our case), and their task is to produce a marginally uniform bit each so that the outputs by Alice and Bob agree with the maximum probability. Although both NICD and the collision complexity are closely related to the maximum correlation and hypercontractivity, their exact relationship is unclear. On the other hand, the relationship between hypercontractivity and maximum correlation is evident in many of the works cited above. We discuss the relationship between these two notions and collision complexity in Subsection 7.4 and the appendix.

The direct precursor to this work is a work of Gavinsky, Ito and Wang [8], which, to our knowledge, was the first that studied different "forms of shared randomness" in communication complexity.

3 Preliminaries

Throughout the paper, the base of logarithm is two unless stated otherwise. For $x, y \in \{0,1\}^n$, by $x \cdot y$ we mean the inner product of x, y as elements of $\mathbb{F}_2^n$, i.e. $\sum_{i=1}^n x_i y_i \bmod 2$.

Recall from the introduction that by a source of shared randomness ρ (also sometimes referred to as a bipartite distributions) we mean a distribution over a set $\Sigma = U \times V$ and for $X \subseteq \Sigma$ the uniform distribution over a set X is denoted by $\mathcal{U}_X$. Also for any distribution ρ on $U \times V$ we denote the marginals of ρ on U and V by ρ_U and ρ_V, respectively.

The main operation on sources of shared randomness is the tensor product.

Definition 3.1. Let ρ_1 and ρ_2 be two distributions over $U_1 \times V_1$ and $U_2 \times V_2$ respectively. We define a new source of shared randomness $\rho_1 \otimes \rho_2$ over $(U_1 \times U_2) \times (V_1 \times V_2)$ by setting

$$(\rho_1 \otimes \rho_2)\left((x,x'),(y,y')\right) = \rho_1(x,y) \cdot \rho_2(x',y')$$

for any $(x,y) \in U_1 \times V_1$ and $(x',y') \in U_2 \times V_2$.

Maximum Correlation. Let ρ be a distribution on $U \times V$. The spaces U and V equipped with the measures ρ_U and ρ_V form two probability spaces which turns the vector space of real-valued functions over U and V into a Hilbert space via $\|f\|_2^2 = \mathbf{E}_{u \sim \rho_A} f(u)^2$ and $\|g\|_2^2 = \mathbf{E}_{v \sim \rho_V} g(v)^2$.

Given the above, we can define the maximum correlation of ρ as follows.

Definition 3.2 (Maximum correlation)**.** Given a distribution ρ on $U \times V$ the maximum correlation of ρ is defined

$$\mathrm{Cor}(\rho) = \sup_{f,g} \mathop{\mathbf{E}}_{(u,v)\sim\rho} [f(u)g(v)]$$

where the supremum is taken over functions satisfying $\mathbf{E}_{u \sim \rho_U} f(u) = \mathbf{E}_{v \sim \rho_V} f(v) = 0$ and $\|f\|_2 = \|g\|_2 = 1$.

From the above definition it is clear that $\mathrm{Cor}(\rho_{perf}) = 1$, which is clearly the largest the maximum correlation can be, because one can take $f = g$ for any $\|f\|_2 = 1$. On the other hand for any product distribution, such as ρ_{priv}, the maximum correlation is zero because in that case by the independence we have $\mathbf{E}_{(u,v)} f(u)f(v) = \mathbf{E}_u f(u) \mathbf{E}_v f(v) = 0$.

The following follows easily from the definition.

Lemma 3.3. *Let ρ be a distribution on $U \times V$ and $f\colon U \to [0,1]$ and $g\colon V \to [0,1]$. Then,*

$$\mathop{\mathbf{E}}_{(u,v)\sim\rho} f(u)g(v) \leq \mathop{\mathbf{E}}_{u} f(u) \mathop{\mathbf{E}}_{v} f(v) + \mathrm{Cor}(\rho) \cdot \sqrt{\mathrm{Var}(f)\,\mathrm{Var}(g)}.$$

Proof. Let $a = \mathbf{E}_{u\sim\rho_U} f(u)$ and $b = \mathbf{E}_{v\sim\rho_V} g(v)$. Note that $\mathbf{E}_{(u,v)\sim\rho} f(u)g(v)$ can be written as $ab + \mathbf{E}_{(u,v)\sim\rho} (f(u)-a)\cdot(g(v)-b)$. The lemma follows from noting that $(f-a)/\sqrt{\mathrm{Var}(f)}$ and $(g-b)/\sqrt{\mathrm{Var}(g)}$ have mean 0 and variance 1 and the definition of the maximum correlation. □

The main attractive feature of maximum correlation is its simple behaviour under *tensor product* of different distributions.[4]

Lemma 3.4 (Witsenhausen [21])**.** *For $i \in [n]$, let ρ_i be a probability distribution on $U_i \times V_i$. Then it holds that* $\mathrm{Cor}(\rho_1 \otimes \cdots \otimes \rho_n) = \max_{i\in[n]} \mathrm{Cor}(\rho_i)$.

Proof. It is easy to see from the definition that maximum correlation of a distribution ρ is just the second largest singular value of the matrix $A_{u,v} = \frac{\rho(u,v)}{\sqrt{\rho(u)\rho(v)}}$. The lemma then follows by noting that the singular values of the tensor product of two matrices are given by the pairwise products of the singular values of the original matrices. □

3.1. Information theory

Here we give a brief account of some of the concepts from Information Theory. For a more detailed introduction containing the omitted proofs, one can consult [6].

For a random variables X over a domain $\mathcal{X}$, its entropy is defined as

$$H(X) = \sum_{x\in\mathcal{X}} \Pr[X = x] \cdot \log\left(\frac{1}{\Pr[X = x]}\right).$$

The entropy of a random variable X conditioned on Y is the average (according to Y) of entropies of random variables $X|Y = y$. More concisely, we have $H(X|Y) = H(XY) - H(Y)$.

Given random variables X and Y we define their mutual information via

$$I(X;Y) = H(X) + H(Y) - H(XY) = H(X) - H(X|Y) = H(Y) - H(Y|X).$$

The main fact that we shall need regarding the mutual information is the following.

Fact 3.5 (Chain rule)**.** For any random variables X, Y, Z, W we have

$$I(X;YZ|W) = I(X;Y|W) + I(X;Z|Y,W).$$

[4] Collision complexity itself also satisfies a nice property under tensor product (though this is essentially by definition and not due to a non-trivial underlying mathematical phenomenon), as explained in Fact 1.7 (iii) and (iv); but these are in general too weak to be of much help.

Finally, we will need a simple lemma which we use in our analysis.

Lemma 3.6. *Let X be a random variable uniformly distributed over the domain $\mathcal{X}$ and let P be an event. Then,*

$$H(X \mid P) \geq H(X) - \log\left(\frac{1}{\Pr[P]}\right)$$

We note that the above lemma does not hold in the current form when X is not uniform. To see this, let X_1, X_2 be uniform independent random variables over $\mathcal{X}_1, \mathcal{X}_2$ respectively. Let $i \in \{1, 2\}$ be also uniform and let $X = X_i$. Taking P to be the event $i = 1$ and assuming that $|\mathcal{X}_1| \ll |\mathcal{X}_2|$ we see that in this case $H(X|P)$ will be much less than $H(X) - 1$ which proves the necessity of uniformity assumption.

Proof of Lemma 3.6. We have

$$\begin{aligned}
H(X|P) &= \sum_{x \in \mathcal{X}} \Pr[X = x|P] \log\left(\frac{1}{\Pr[X = x|P]}\right) = \sum_{x \in \mathcal{X}} \Pr[X = x|P] \cdot \log\left(\frac{\Pr[P]}{\Pr[X = x \wedge P]}\right) \\
&= \log(\Pr[P]) + \sum_{x \in \mathcal{X}} \Pr[X = x|P] \cdot \log\left(\frac{1}{\Pr[X = x \wedge P]}\right) \\
&\geq \log(\Pr[P]) + \sum_{x \in \mathcal{X}} \Pr[X = x|P] \cdot \log\left(\frac{1}{\Pr[X = x]}\right) \\
&= \log(\Pr[P]) + \sum_{x \in \mathcal{X}} \Pr[X = x|P] \cdot \log|\mathcal{X}| \\
&= \log(\Pr[P] \cdot |\mathcal{X}|),
\end{aligned}$$

where the latter is clearly equal to $H(X) - \log(\frac{1}{\Pr[P]})$. □

3.2. Communication complexity

We assume some basic familiarity with communication complexity [14]. For a partial function f from $\{0,1\}^n \times \{0,1\}^n$ to E, we denote by $SMP(f)$ the SMP communication complexity of f without shared randomness with worst-case error probability (over the private coins of Alice and Bob) at most $1/3$.

For $n \geq 1$, let $IP_n \colon \{0,1\}^n \times \{0,1\}^n \to \{0,1\}$ be the inner product function modulo 2. Chor and Goldreich [5] showed that any communication protocol which answers IP_n with error probability $1/3$ on average, even in the two-way communication model, must have communication cost at least $n/2 - o(n)$. The following is an immediate corollary of this.

Corollary 3.7. *For $n \geq 1$, the communication complexity of IP_n in the SMP model with public randomness is in $\Omega(n)$.*

We will use the following fact originally proved by Newman [18] and refined by Kushilevitz and Nisan [14, Theorem 3.14]. The statement in [14] assumes that the function takes a Boolean value, but the proof does not use this assumption.

Lemma 3.8. *Let $0 < \varepsilon < \varepsilon' < 1/2$ be constants. Any SMP protocol with public randomness for a partial function from $\{0,1\}^n \times \{0,1\}^n$ to E with error probability at most ε can be converted to*

one which uses only $\log n + C$ *bits of public randomness with error probability at most* ε' *without changing the communication cost, where* $C = C(\varepsilon, \varepsilon') > 0$ *is a constant depending only on* ε *and* ε'.

4 Usefulness of all non-product sources of shared randomness

In this section, we prove Theorem 1.5 showing an easy but nontrivial use of imperfect shared randomness in an SMP protocol for equality function.

As a consequence of this theorem, we see that the lower bound claimed in the second part of our main theorem (Theorem 1.8) does not hold for an arbitrary family of functions $\{g_n\}$. Thus a good choice of a function family was necessary there.

Beyond its bearing on Theorem 1.8, the result is interesting on its own as it explicitly shows that in some settings *any form of public shared randomness* (i.e. one coming from a non-product source ρ) can have a significant impact on the asymptotic communication cost.

Proof of Theorem 1.5. Because ρ is not a product distribution, there exist subsets $\Lambda_a \subseteq U$ and $\Lambda_b \subseteq V$ such that

$$\gamma = \Pr_{(u,v)\sim\rho}[u \in \Lambda_a \wedge v \in \Lambda_b] \neq \Pr_{u\sim\rho_A}[u \in \Lambda_a] \Pr_{v\sim\rho_B}[v \in \Lambda_b] = \gamma'. \tag{3}$$

A protocol for solving the equality is as follows. Alice and Bob share 2^n copies of shared randomness; label the 2^n i.i.d. pairs of registers containing the shared randomness as $(u_x, v_x) \sim \rho$ for $x \in \{0,1\}^n$. Alice defines $\alpha = 1$ if $u_x \in \Lambda_a$ and $\alpha = 0$ otherwise, and sends α to the referee. Similarly, Bob defines $\beta = 1$ if $v_y \in \Lambda_b$ and $\beta = 0$ otherwise, and sends β to the referee. The referee checks whether $\alpha = \beta = 1$ or not. If $x = y$, this happens with probability γ, and otherwise it happens with probability γ'. The referee can tell which is the case with error probability at most $1/3$ by repeating this protocol for $t = O(|\gamma - \gamma'|^{-2})$ which is a constant independent of n. □

5 Collision complexity

In the introduction we gave an informal definition and briefly mentioned some of the basic properties of collision complexity. In this section, we look at this notion in more detail and provide a more formal definition. We also give an alternate and more analytic characterisation of collision complexity in terms of a different measure of quality of shared randomness called *agreement complexity*. Beside providing a different perspective on collision complexity, this alternate characterisation will be also useful technically at various places later.

We start our work toward the formal definition of collision complexity by defining a collision protocol.

Definition 5.1 (Collision protocol)**.** Let ρ be a probability distribution on $U \times V$. A p-collision protocol [5] for ρ with domain size n is determined by two functions, $A : U^\ell \to \mathcal{P}[n]$ and $B : V^\ell \to \mathcal{P}[n]$ for some $\ell \in \mathbb{N}$, such that

$$\forall i \in [n] : \qquad \Pr_{(\vec{u},\vec{v})\sim\rho^{\otimes\ell}}[(\vec{u},\vec{v}) : \ i \in A(\vec{u}) \cap B(\vec{v})] \geq p. \tag{4}$$

[5] Think of p as a decreasing function of n: say, $p = 1/n$.

It is sometimes convenient to think of Alice's and Bob's functions A and B as randomised mappings, and that is allowed: we always assume that ρ is non-degenerate, so the players can use it to generate private randomness (it doesn't have to be hidden from the other party, as the players are always cooperating).

The *complexity* of a collision protocol $\{A, B\}$ is given by the maximum output size:

$$\mathrm{col}_\rho(n, p, \{A, B\}) \stackrel{\mathrm{def}}{=} \max\{|A(\vec{u})|, |B(\vec{v})| \,:\, (\vec{u}, \vec{v}) \in \mathit{supp}(\rho)\}\,. \tag{5}$$

Definition 5.2 (Collision complexity)**.** Let $n \in \mathbb{N}$ and $p = p(n) \in [0, 1]$ be a probability parameter possibly depending on n. The p-collision complexity of a source ρ (over a domain of size n) is the minimum worst case output size necessary for any protocol achieving a collision probability of p for all $i \in [n]$. In other words,

$$\mathrm{col}_\rho(n, p) = \inf_{\ell, A, B} \mathrm{col}_\rho(n, p, \{A, B\}),$$

where the infimum is over all protocols (ℓ, A, B) satisfying (4).

The main parameter setting of interest for us here will be $p = 1/n$ and as such what follows we shall let

$$\mathrm{col}_\rho(n) \stackrel{\mathrm{def}}{=} \mathrm{col}_\rho(n, 1/n). \tag{6}$$

Note that the collision complexity of a non-degenerate distribution can be at most $O(\sqrt{n})$ (achievable using private randomness via the birthday paradox). Similarly, the rest of the properties from Fact 1.7 easily follow from the definition.

Our second measure of correlation is *agreement complexity*. It is closely related to collision complexity, but sometimes will be more convenient to work with.

Definition 5.3 (Agreement complexity)**.** Let ρ be a probability distribution on $U \times V$.

An *agreement protocol* for ρ is determined by $\ell \in \mathbb{N}$ and a pair of functions $f\colon U^\ell \to [0, 1]$ and $g\colon V^\ell \to [0, 1]$. The *cost* of this agreement protocol is $\mathbf{E}[f(u_1, \ldots, u_\ell) + g(v_1, \ldots, v_\ell)]$, and the *success probability* of this protocol is $\mathbf{E}[f(u_1, \ldots, u_\ell)g(v_1, \ldots, v_\ell)]$, where $(u_i, v_i) \sim \rho$ independently for all $i \in [\ell]$.

The *agreement complexity* of ρ at success probability p, denoted by $\mathrm{agr}_\rho(p)$, is the infimum of the cost of an agreement protocol for ρ with success probability at least p.

Here Alice and Bob output one bit each (as opposed to a subset of $[n]$ in the case of collision complexity). The value $f(u_1, \ldots, u_\ell)$ is the probability that Alice outputs "1", given her part of shared randomness, and similarly for $g(v_1, \ldots, v_\ell)$. The players' task is to output "1" simultaneously with probability at least p, while minimising the sum of the probabilities that each party outputs "1" – the corresponding infimum is the agreement complexity.

5.1. Equivalence of collision and agreement complexities and further properties

The first lemma provides some basic parameter trade-offs.

Lemma 5.4. *For any positive integers m, n and $p \in (0, 1)$ we have:*

$$\mathrm{col}_\rho(n, 1 - (1 - p)^m) \le m\,\mathrm{col}_\rho(n, p), \qquad \mathrm{col}_\rho(mn, p) \le m\,\mathrm{col}_\rho(n, p). \tag{7}$$

Proof. The first inequality follows by repeating a collision protocol m times independently and outputting the union of the results. For the second one, repeat the collision protocol m times independently to obtain $A_1, \dots, A_m \subseteq [n]$ and $B_1, \dots, B_m \subseteq [n]$, and output $A = \{n(i-1)+j \colon i \in [m], j \in A_j\}$ and $B = \{n(i-1)+j \colon i \in [m], j \in B_j\}$. This gives a collision protocol with domain size mn, collision probability p, and output size at most $m \operatorname{col}_\rho(n,p)$. □

As we mentioned collision and agreement complexities are intimately related. This is captured by the following lemma.

Lemma 5.5. *For a bipartite distribution ρ, $n \geq 1$, and $0 < p < 1$, it holds that $\operatorname{col}_\rho(n,p) = \Theta(\max\{1, n \operatorname{agr}_\rho(p)\})$, where the constant in the asymptotic notation does not depend on ρ, n, or p.*

The proof is based on the following idea: If $i \in [n]$, a collision protocol (ℓ, A, B) of domain size n can be converted to an agreement protocol where Alice outputs 1 if $i \in A$ and Bob outputs 1 if $i \in B$. By choosing the "best" value of i, we obtain the claimed upper bound on the agreement complexity. For the conversion in the opposite direction, Alice and Bob repeat the agreement protocol n times in parallel to decide whether their output should contain each $i \in [n]$. By the Chernoff bound, this provides the necessary upper bound on the collision complexity after an application of Lemma 5.4. The formal proof is as follows:

Proof. Let (ℓ, A, B) be a collision protocol for ρ with domain size n, output size $\operatorname{col}_\rho(n,p)$, and collision probability at least p. For $i \in [n]$, let $f_i(u) = \Pr[i \in A(u)]$ for $\vec{u} \in U^\ell$ and let $g_i(\vec{v}) = \Pr[i \in B(\vec{v})]$ for $\vec{v} \in V^\ell$. Because the collision probability of (ℓ, A, B) is at least p, it holds that $\mathbf{E}[f_i(\vec{u}) g_i(\vec{v})] \geq p$ for every $i \in [n]$. Because The output size being is $\operatorname{col}_\rho(n)$, it holds that $|A(\vec{u})| \leq \operatorname{col}_\rho(n,p)$ and $|B(\vec{v})| \leq \operatorname{col}_\rho(n,p)$ for all $\vec{u} \in U^\ell$ and $\vec{v} \in V^\ell$, and in particular, it implies that

$$\mathbf{E}[|A(\vec{u})| + |B(\vec{v})|] = \sum_{i \in [n]} \mathbf{E}[f_i(\vec{u}) + g_i(\vec{v})] \leq 2 \operatorname{col}_\rho(n,p).$$

Therefore, there exists $i^* \in [n]$ such that

$$\mathbf{E}[f_{i^*}(\vec{u}) + g_{i^*}(\vec{v})] \leq \frac{2 \operatorname{col}_\rho(n,p)}{n}.$$

Then the pair (f_{i^*}, g_{i^*}) is an agreement protocol with success probability at least p and cost at most $2\operatorname{col}_\rho(n,p)/n$. Therefore, we have that $n \operatorname{agr}_\rho(p) \leq 2 \operatorname{col}_\rho(n,p)$.

Next we prove that $\operatorname{col}_\rho(n,p) \leq 9n \operatorname{agr}_\rho(p) + 48$. Because $\operatorname{agr}_\rho(p) \geq 2p$, this inequality is trivial if $p \geq 1/2$. For the rest of the proof, assume $p < 1/2$, and we will prove the inequality in two steps:

$$\operatorname{col}_\rho(n, p/2) \leq 3n \operatorname{agr}_\rho(p) + 16, \tag{8}$$

$$\operatorname{col}_\rho(n,p) \leq 3 \operatorname{col}_\rho(n, p/2). \tag{9}$$

Let $K > \operatorname{agr}_\rho(p)$. By the definition of the agreement complexity, there exists an agreement protocol (ℓ, f, g) for ρ with success probability at least p and cost less than c. Let $a = \mathbf{E}_{\vec{u} \sim \rho_{U^\ell}}[f(\vec{u})]$ and $b = \mathbf{E}_{\vec{v} \sim \rho_{V^\ell}}[g(\vec{v})]$, and we have that $a < K$ and $b < K$. Let $T = 3nK + 16$. Consider the following collision protocol. Alice interprets her random input as $(\vec{u}_1, \dots, \vec{u}_n) \in (U^\ell)^n$, and Bob interprets his random input as $(\vec{v}_1, \dots, \vec{v}_n) \in (V^\ell)^n$, so that $(\vec{u}_i, \vec{v}_i)$ is distributed according to $\rho^{\otimes \ell}$ for each $i \in [n]$ and n pairs $(\vec{u}_1, \vec{v}_1), \dots, (\vec{u}_n, \vec{v}_n)$ are mutually independent. Alice constructs

sets $\tilde{A}, A \subseteq [n]$ as follows. For each $i \in [n]$, i belongs to $\tilde{A}$ with probability $f(\vec{u}_i)$, independently of everything else. If $|\tilde{A}| \leq T$, then she defines $A = \tilde{A}$; otherwise $A = \varnothing$. She outputs set A. Bob constructs sets $\tilde{B}$ and B in the analogous way, and outputs B. It is clear that the output size of this collision protocol is at most t.

Let $i \in [n]$. By the Chernoff bound, it holds that

$$\begin{aligned}\Pr[A \neq \tilde{A} \mid i \in \tilde{A} \cap \tilde{B}] &= \Pr[|\tilde{A}| > T \mid i \in \tilde{A} \cap \tilde{B}] \\ &= \Pr[|\tilde{A} \setminus \{i\}| > t-1 \mid i \in \tilde{A} \cap \tilde{B}] \\ &\leq \frac{1}{e^{a(n-1)}} \cdot \left(\frac{ea(n-1)}{T-1}\right)^{T-1} \\ &< \left(\frac{e}{3}\right)^{T-1} < \frac{1}{4},\end{aligned}$$

where the last inequality follows from $(e/3)^{15} < 1/4$. Similarly, it holds that $\Pr[B \neq \tilde{B} \mid i \in \tilde{A} \cap \tilde{B}] < 1/4$. By union bound, we have that

$$\Pr[A = \tilde{A} \wedge B = \tilde{B} \mid i \in \tilde{A} \cap \tilde{B}] > \frac{1}{2}.$$

Therefore,

$$\Pr[i \in A \cap B] = \Pr[i \in \tilde{A} \cap \tilde{B}] \Pr[A = \tilde{A} \wedge B = \tilde{B} \mid i \in \tilde{A} \cap \tilde{B}] > \frac{p}{2},$$

meaning that the collision probability of this collision protocol is greater than $p/2$. This implies that $\mathrm{col}_\rho(n, p/2) \leq 3nK + 16$. Because this holds with any $K > \mathrm{agr}_\rho(p)$, inequality (8) follows.

Inequality (9) follows from (7) by noting that $1 - (1 - p/2)^3 > p$ because $0 < p < 1/2$. □

6 Simulation of SMP_{perf} protocols in SMP_ρ

Given the technical tools we developed in the previous section we are ready to prove the first (easy) part of Theorem 1.8.

Proposition 6.1. *Let ρ be a bipartite distribution. Then for any (possibly partial) function f over $\{0,1\}^n \times \{0,1\}^n$ we have*

$$SMP_\rho(f) = O(\mathrm{col}_\rho(n)\,(SMP_{perf}(f) + \log n)\,).$$

To prove the lemma, we first need a symmetrisation claim which shows that at a small multiplicative cost one can always turn a collision protocol into one which treats all elements $i \in [n]$ the same way. More formally, we have the following:

Claim 6.2. Let $0 < s < 1$. For a bipartite distribution ρ and $n \geq 1$, there exists a collision protocol (ℓ, A, B) with domain size n and output size at most $O(\mathrm{col}_\rho(n))$ that satisfies the following two properties:

(a) $\Pr_{(u,v)\sim\rho^{\otimes\ell}}[A(u) \cap B(v) = \varnothing] \leq s$.

(b) Conditioned on the event $A(u) \cap B(v) \neq \varnothing$, selecting one element from $A(u) \cap B(v)$ uniformly at random produces the uniform distribution over $[n]$.

The constant in the asymptotic notation depends on s, the failure probability which is a small but fixed constant, however it does not depend on ρ or n.

Proof. We first prove the case where $s \geq 1/e+1/2$. Consider the collision protocol used in the proof of inequality (8) in Lemma 5.5 with $p = 1/n$. This protocol satisfies condition (b) by symmetry. Its output size is at most $3n \operatorname{agr}_\rho(1/n)+16$, which is in $O(\operatorname{col}_\rho(n))$ by Lemma 5.5. By the Chernoff bound, it holds that

$$\Pr[A \neq \tilde{A}] = \Pr[|\tilde{A}| > T] \leq \frac{1}{e^{an}} \cdot \left(\frac{ean}{T}\right)^T < \left(\frac{e}{3}\right)^T < \frac{1}{4},$$

and similarly $\Pr[B \neq \tilde{B}] < 1/4$. By union bound, it holds that

$$\begin{aligned}\Pr[A \cap B = \varnothing] &\leq \Pr[\tilde{A} \cap \tilde{B} = \varnothing] + \Pr[A \neq \tilde{A}] + \Pr[B \neq \tilde{B}] \\ &< \left(1 - \frac{1}{n}\right)^n + \frac{1}{4} + \frac{1}{4} \\ &< \frac{1}{e} + \frac{1}{2}.\end{aligned}$$

This proves the case where $s \geq 1/e + 1/2$.

To prove the general case, repeat this protocol for $\lceil \log s / \log(1/e + 1/2) \rceil$ times and output the union of the results. □

Proof of Proposition 6.1. Using randomness reduction, i.e. Lemma 3.8, we know that there exists an SMP protocol for f using at most $\log n + C$ bits of public perfect randomness achieving with error probability at most $2/5$. Fix this protocol and set $m = \log n + C$ and $\epsilon = 2/5$ and $s = (1 - 2\varepsilon)/(4 - 4\varepsilon)$. Let $\alpha(x, r)$ to be the message which Alice sends given input x and public (perfect) random string $r \in 2^{[m]}$, and let $\beta(y, r)$ be the message which Bob sends given input is y and randomness r. Let $\gamma(a, b) \in E$ be the referee's output given that the message from Alice is a and the message from Bob is b.

Consider the following SMP protocol for f with shared distribution ρ.

1. Alice and Bob run the collision protocol in Claim 6.2 with domain size 2^m and parameter s specified above. Identify the domain $[2^m]$ with the set of m-bit strings, and let $A, B \subseteq \{0,1\}^m$ be the sets that Alice and Bob obtain as a result of running this collision protocol. Alice sends the referee each string r in A together with $\alpha(x, r)$, and Bob sends the referee each string r in B together with $\beta(y, r)$.
2. The referee outputs an arbitrary answer in E if $A \cap B = \varnothing$. Otherwise, he chooses an element $r \in A \cap B$ uniformly at random, and he outputs $\gamma(\alpha(x, r), \beta(y, r))$ (which he can do, because the values $\alpha(x, r)$ and $\beta(y, r)$ have been sent by Alice and Bob, respectively).

The complexity of the above protocol is clearly in $O(\operatorname{col}_\rho(2^m)(SMP_{perf} + m))$. It is clear that the protocol succeeds with probability $(1 - s)(1 - \varepsilon) = (3 - 2\varepsilon)/4 > 1/2$. By amplification, we are done. □

7 Characterising $SMP_\rho(f)$ via collision complexity: the lower bound

In the previous section we proved the first part of Theorem 1.8. In this section, we complete what we started there by proving the second (and the harder) part of the theorem. The main ingredient of the proof is a lower bound on the communication complexity of $GAPIP_n$ functions (introduced in Definition 1.9) which is the content of Theorem 7.1.

Theorem 7.1. *For any bipartite distribution ρ, it holds that $SMP_\rho(GAPIP_n) = \Omega(\mathrm{col}_\rho(n))$, where the constants in the asymptotic notations do not depend on ρ or n.*

The proof of the above Theorem is rather involved and requires several steps. We shall give an exposition of the main ideas behind the proof in the next subsection, and then carry out the proof in details in the remaining three subsections. Before delving into that, however, let us observe that this Theorem 7.1 together with two easy propositions imply the second half of Theorem 1.8.

Proposition 7.2. *For a bipartite distribution ρ, it holds that $SMP_\rho(GAPIP_n) = O(\mathrm{col}_\rho(n) \log n)$.*

Proof. Note that $GAPIP_n$ has a straightforward SMP protocol with public randomness with cost $O(\log n)$ and error probability at most $1/3$: Alice and Bob choose a common index $i \in [n]$ uniformly at random, Alice sends x_i to the referee, and Bob sends y_i to the referee. The referee simply answers $x_i \cdot y_i$. Because this protocol uses $\log n$ bits of public randomness, Lemma 6.1 (i) implies the claim. □

Proposition 7.3. *For a bipartite distribution ρ, it holds that $SMP_\rho(GAPIP_n) = \Omega(\log n)$.*

Proof. Note that IP_m can be reduced to $GAPIP_{n,m}$ by repeating the input vector n times, and in particular $IP_{8 \log n}$ can be reduced to $GAPIP_n$. By Corollary 3.7, this implies the claim. □

From the above we get the following corollary which was our main goal.

Corollary 7.4 (Second half of Theorem 1.8)**.** *There exists a family of partial functions f_N from $\{0,1\}^N \times \{0,1\}^N$ to $\{0,1\}$ such that for any bipartite distribution ρ, it holds that*

$$SMP_\rho(f_N) = \begin{cases} O\left(\mathrm{col}_\rho\left(\frac{N}{\log N}\right) \log N\right) & (\textit{upper bound}) \\ \Omega\left(\max\left\{\log N, \mathrm{col}_\rho\left(\frac{N}{\log N}\right)\right\}\right) & (\textit{lower bound}) \end{cases},$$

where the constants in the asymptotic notations do not depend on ρ or N.

The reason for appearance of $\frac{N}{\log N}$ terms in the above statement is that the input size to $GAPIP_n$ is $N = nm = 8n \log n$ which essentially is the main source of the poly-logarithmic gap between the upper and lower bounds in the above. However, note that the upper and lower bounds in Corollary 7.4 are still quite close, because thanks to inequality (7) we have $\mathrm{col}_\rho(N/\log N) = \Omega(\mathrm{col}_\rho(N)/\log N)$—which means that the result is tight up to a factor of $O(\log^2 N)$ for every bipartite distribution ρ.

7.1. The proof outline of Theorem 7.1

The proof of Theorem 7.1 goes roughly as follows. Consider a hypothetical protocol for $GAPIP_n$ with complexity $o(\mathrm{col}_\rho(n))$. We will consider the behaviour of the protocol under the uniformly random input distribution—ignoring the promise on the input of $GAPIP_n$ for the moment. Because

of the low communication complexity of the protocol we can argue (using a "rounding argument" which extracts a collision protocol from a $GAPIP_n$ protocol by looking at the influential coordinates) that for a typical $i \in [n]$ the referee can guess the value of $x_i \cdot y_i$ correctly only with probability at most $1/2 + O(1/\text{poly}(n))$ where the poly(n) here is a polynomial with *super-linear growth*.

Next, we will use the repeat the above argument inductively $\Omega(n)$ times to conclude that there exists a set $L \subseteq [n]$ of size at least $2n/3$, such that the referee can distinguish the case when $x_i \cdot y_i = 0$ for all $i \in L$ from the case when $x_i \cdot y_i = 1$ for all $i \in L$ only with probability $1/2 + o(1)$, where here we crucially rely on the fact the bias of the protocol in the previous step was only $o(1/n)$. Finally, if we define the input distribution μ to be uniform over $\{(x, y) \colon \forall i \in L : x_i = y_i\} \cup \{(x, y) \colon \forall i \in L : x_i \neq y_i\}$, then each element in the support of μ is a valid input to $GAPIP_n$, but the protocol under consideration makes a mistake with respect to μ with large probability which is in contradiction with the initial hypothesis.

The above outline essentially contains a full description of our argument except for one subtle but important technical point: It turns out that as the hybrid argument progresses, turning the distribution of (x_i, y_i) from uniform to $x_i \cdot y_i = 0$ (say) for many coordinates i, the players may in principle be able to use the extra shared randomness offered by conditioning on the previous rounds to their advantage. All of this essentially amounts to the fact, that in the midst of the hybrid argument the players instead of just having access to ρ, instead have access to the more powerful source $\rho \otimes \sigma_0$ (or $\rho \otimes \sigma_1$)—where σ_b denotes the uniform distribution on $(x^*, y^*) \in \mathbb{F}_2^m \times \mathbb{F}_2^m$ conditioned on $x^* \cdot y^* = b$. In fact, we do not know any way to show that the players cannot take advantage of the newly introduced shared randomness caused by conditioning, rather we use some ideas related to maximum correlation to show that, for our specific choice of parameter $m = 8 \log n$, one has

$$\text{col}_{\rho \otimes \sigma_0}(n) = \Omega(\text{col}_\rho(n)), \qquad \text{col}_{\rho \otimes \sigma_1}(n) = \Omega(\text{col}_\rho(n)), \tag{10}$$

which suffices for our application.

We note that similar issues to the one above, i.e. the emergence of additional undesired correlations between parties as a result of conditioning, occur in many different contexts in the theoretical computer science literature, for example in the context of parallel repetition [10, 20] and also in many other places in communication complexity [3, 11]. However, the method we use here to handle this issue is quite different from the prior work, and hence, could potentially be of interest for those applications as well.

7.2. Hardness of RIP_n: the inductive step

In our analysis, it is convenient to use a slight variant of the standard SMP model, which we call the *pseudo-SMP* model, with the following properties:

- The referee can receive his own portion of input.
- If shared randomness is available, the referee can see both Alice's and Bob's part of the shared distribution.

The crucial condition here is the latter, i.e. the fact that Alice and Bob's shared randomness (but not inputs) is visible to the referee. Note that in SMP, the players and the referee cooperate in order to solve the problem, giving the referee the above extra power only makes the model stronger; thus,

any lower-bound established for pseudo-SMP immediately implies a corresponding lower bound for the usual SMP model (possibly with some input the referee).[6]

For our lower bound it is natural to introduce some auxiliary communication tasks. First consider the following general "guessing" information theoretic task, which is a two-player variant of a problem considered previously by Gavinsky, Kempe, Regev, and de Wolf [9]:

Definition 7.5 (Random access problem)**.** In $RA_{n,\delta}$ problem, Alice receives $X \in \{0,1\}^n$, Bob receives $Y \in \{0,1\}^n$, and the referee receives $I \in [n]$. A pseudo-SMP protocol solves $RA_{n,\delta}$ problem successfully if the referee's output $Z \in \{0,1\}^2 \cup \{\perp\}$ satisfies

$$\Pr_{X,Y\sim\mathcal{U}}[Z = (X_i, Y_i) \mid I = i,\ Z \neq \perp] \geq 1 - \delta, \qquad \forall\, i \in [n].$$

Next we introduce the Random Access Inner Product Problem, which is slightly less information theoretic, and closer to $GAPIP_n$.

Definition 7.6 (Random access inner product problem)**.** Let $n, m \in \mathbb{N}$ and suppose $x_i, y_i \in \mathbb{F}_2^m$ for $i \in [n]$ Denote $x = (x_1, \ldots, x_n)$ and $y = (y_1, \ldots, y_n)$. Then $RIP_{n,m}(x, y, i) \stackrel{\text{def}}{=} x_i \cdot y_i$.

We will write RIP_n to address $RIP_{n,8\log n}$ and let $I = i$ denote the event that the third input to $RIP_n(\cdot,\cdot,I)$ is fixed to i.

The main purpose of this subsection is to prove the next lemma, which states that RIP_n is a hard problem; this is essentially all we need in the next subsection to run our inductive argument for lower bounding $SMP_\rho(GAPIP_n)$ there.

Lemma 7.7. *Let $\mathcal{P}$ be a pseudo-SMP protocol for RIP_n using source ρ and communication cost $CC_\mathcal{P}$ such that $CC_\mathcal{P} + \log n \leq 41 \cdot \mathrm{col}_\rho(n)$. Then there exists $i \in [n]$ such that*

$$\gamma_i \leq \frac{1}{2} + O\left(\frac{CC_\mathcal{P} + \log n}{n\,\mathrm{col}_\rho(n)}\right) + o\left(\frac{1}{n}\right),^{7} \tag{11}$$

where γ_i is the probability of success of $\mathcal{P}$ given $I = i$ and uniformly random inputs $x, y \in (\mathbb{F}_2^m)^n$.

To understand the statement of this lemma, consider the case where $\mathrm{col}_\rho(n) = \omega(\log n)$. Then the lemma claims that if the cost $CC_\mathcal{P}$ of a communication protocol $\mathcal{P}$ for RIP_n is too small, i.e., $CC_\mathcal{P} = o(\mathrm{col}_\rho(n))$, then there is a coordinate $i \in [n]$ such that $\mathcal{P}$ answers $x_i \cdot y_i$ correctly with probability at most $1/2 + o(1/n)$.

To prove Lemma 7.7, we first show that $RA_{n,\delta}$ is a hard problem and then reduce $RA_{n,\delta}$ to RIP_n, thus establishing that the latter is also hard. More precisely, we prove the following.

Lemma 7.8. *There exists a constant $\delta > 0$ such that the following holds. Let $\mathcal{P}$ be a pseudo-SMP protocol for $RA_{n,\delta}$ using source ρ and communication cost $CC_\mathcal{P}$ such that $36(CC_\mathcal{P} + \log n) \leq \mathrm{col}_\rho(n)$. Then there exists $i \in [n]$ such that the probability of referee outputting anything beside $\perp$ satisfies*

$$\Pr_{x,y}[\mathcal{P}(x, y, i) \neq \perp] = O\left(\frac{CC_\mathcal{P} + \log n}{n\,\mathrm{col}_\rho(n)}\right) + o\left(\frac{1}{n}\right),$$

where $x, y \in \{0,1\}^n$ are chosen uniformly at random.

[6] The reason we do this here is that to prove the lower bound in Lemma 7.7 (even for just basic SMP), we seem to need a lower bound for $RA_{n,\delta}$ against the slightly stronger pseudo-SMP model.

[7] The implied constants in $O(\cdot)$ and $o(\cdot)$ here are universal and, in particular, do not depend on ρ or n.

Proof. The main idea is to use the protocol $\mathcal{P}$ for $RA_{n,\delta}$ to produce a good collision protocol (over domain of size n) by considering the influential coordinates of $\mathcal{P}$. Following the above idea, it turns out the collision probability of the protocol we obtain will be directly related to probability that the referee did not output $\perp$ in the original $RA_{n,\delta}$ protocol which implies the desired result. The details are as follows.

Let $(R_A, R_B) \sim \rho^{\otimes \ell}$ denote the shared randomness between Alice and Bob. Let M_A and M_B be the random variables representing Alice and Bob's messages to the referee. We also let Z denote the referee's output.

Define random variables L_A and L_B as a function of $R_A = r_A$, $M_A = m_a$, $R_B = r_B$ and $M_B = m_b$ by

$$L_A = \left\{ i \in [n] \colon H(X_i | R_A = r_A, M_A = m_a) < \frac{1}{2} \right\},$$
$$L_B = \left\{ i \in [n] \colon H(Y_i | R_B = r_B, M_B = m_B) < \frac{1}{2} \right\}.$$

In other words, L_A (resp. L_B) consists of the coordinates of Alice's inputs (resp. Bob's inputs) where the referee, who sees both of r_A, m_A (resp. r_B, m_B) by definition of pseudo-SMP model[8] have substantial information about.

Since we are assuming M_A and M_B are small, we expect that typically the sets L_A and L_B are also small. More formally, we show that for any $r_A \in U^\ell$

$$\Pr[|L_A| \geq t \mid R_A = r_A] < \cdot 2^{CC_{\mathcal{P}} - t/2}. \tag{12}$$

where the probability is over a random choice of $x \in \{0,1\}^n$. By the definition of L_A we have

$$\begin{aligned}
&H(X \mid |L_A| \geq t, R_A = r_A) \\
&\leq H(M_A \mid |L_A| \geq t,\, R_A = r_A) + H(X \mid M_A,\, |L_A| \geq t,\, R_A = r_A) \\
&= CC_{\mathcal{P}} + \sum_{i=1}^{n} H(X_i | X_{<i}, M_A,\, |L_A| \geq t,\, R_A = r_A) \\
&\leq CC_{\mathcal{P}} + \sum_{i=1}^{n} H(X_i | M_A,\, |L_A| \geq t,\, R_A = r_A) \\
&\leq CC_{\mathcal{P}} + n - \frac{t}{2}.
\end{aligned}$$

Now we are ready to describe our collision protocol $T = 2C + 4\log n$ By Lemma 3.6, this implies inequality (12).

Because inequality (12) holds for every choice of r_A, it holds without conditioning, i.e.

$$\Pr[|L_A| \geq t] < 2^{CC_{\mathcal{P}} - t/2}.$$

Setting $T = 2C + 4\log n$ and using the symmetric argument for Bob, we obtain that

$$\Pr[|L_A| \geq T \vee |L_B| \geq T] < \frac{2}{n^2}. \tag{13}$$

[8]Recall that in the pseudo-SMP model, the referee sees all the shared randomness as well as the messages from Alice and Bob.

Recall that our goal was to "extract" from $\mathcal{P}$ a *collision protocol* $\mathcal{P}_{col}$. This is given as follows: In $\mathcal{P}_{col}$ Alice and Bob pick random inputs $X, Y \sim \mathcal{U}_{\{0,1\}^n}$ and respective half of the shared randomness $(R_A, R_B) \sim \rho^{\otimes \ell}$, and sample M_A and M_B by simulating the action of the players in $\mathcal{P}$. The players can each locally compute as L_A, L_B as these are a deterministic functions of $M_A R_A$ and $M_B R_B$ respectively. The output of the protocol is $L'_A, L'_B \subseteq [n]$, where

$$L'_A \stackrel{\text{def}}{=} \begin{cases} L_A & |L_A| < T \\ \emptyset & |L_A| \geq T \end{cases}, \qquad L'_B \stackrel{\text{def}}{=} \begin{cases} L_B & |L_B| < T \\ \emptyset & |L_B| \geq T \end{cases}.$$

The analysis of collision protocol. Let $\alpha = \min_{i \in [n]} \Pr[i \in L_A \cap L_B]$. From (13) it follows that $\forall i \in [n] : \Pr[i \in L'_A \cap L'_B] > \alpha - \frac{2}{n^2}$.

This means that

$$\mathrm{col}_\rho\left(n, \alpha - \frac{2}{n^2}\right) \leq T.$$

By inequality (7), this implies that

$$\mathrm{col}_\rho(n) \leq T \cdot \left\lceil \frac{-\ln(1-1/n)}{-\ln(1-\alpha+2/n^2)} \right\rceil \leq T \cdot \left(1 + \frac{2/n}{\alpha - 2/n^2}\right),$$

where the second inequality follows from $-\ln(1-x) \leq 2x$ for $0 < x \leq 1/2$ and $-\ln(1-y) \geq y$ for $y < 1$. If $T \leq \mathrm{col}_\rho(n)$, then this implies that

$$\alpha \leq \frac{4}{n} \cdot \frac{T}{\mathrm{col}_\rho(n)} + \frac{2}{n^2}. \tag{14}$$

Next we need to related α to the probability that the referee outputs $\perp$ in $\mathcal{P}$. Combined with (14) this will finish the proof. Suppose $\delta < \frac{1}{41}$. Since $\mathcal{P}$ is a correct protocol for $RA_{n,\delta}$ and since $h(x) = -x \log x - (1-x)\log(1-x) < 1/6$ for $0 \leq x \leq \delta$, we have

$$H(X_i \mid R_A R_B M_A M_B, I = i, Z \neq \perp) < \frac{1}{6}.$$

Since X_i and $R_B M_B$ are independent conditioned on R_A and M_A, we immediately see that

$$\mathbf{E}_{r_A, m_B}[H(X_i | R_A = r_A, M_A = m_A) \mid I = i \ Z \neq \neg\perp] < \frac{1}{6},$$

and so by Markov's inequality we see that

$$\Pr[i \in L_A \mid I = i, Z \neq \perp] \geq 2/3.$$

Using the similar argument for Bob and the union bound, we obtain that $\Pr[i \in L_A \cap L_B \mid I = i,\ Z \neq \perp] \geq 1/3$. Thus, for all $i \in [n]$ we have

$$\begin{aligned} \Pr[I = i,\ Z \neq \perp] &\leq 3 \cdot \Pr[i \in L_A \cap L_B,\ I = i,\ Z \neq \perp] \\ &\leq \Pr[i \in L_A \cap L_B,\ I = i] \\ &= \Pr[i \in L_A \cap L_B] \cdot \Pr[I = i], \end{aligned}$$

which after dividing by $\Pr[I = i] = \frac{1}{n}$ gives us

$$\Pr[Z \neq \perp \mid I = i] \leq 3 \Pr[i \in L_A \cap L_B] = 3\alpha. \tag{15}$$

Combining (15) and (14) we have

$$\Pr[Z \neq \perp \mid I = i] \leq 3\alpha = O\left(\frac{CC_{\mathcal{P}} + \log n}{n \cdot \mathrm{col}_\rho(n)} + \frac{1}{n^2}\right),$$

which finishes the proof. □

It remains to see that $RA_{n,\delta}$ can be reduced to RIP_n. Here is where the fact that our lower bound for $RA_{n,\delta}$ in Lemma 7.8 is against the stronger pseudo-SMP model comes handy as our reduction from $RA_{n,\delta}$ to RIP_n only produces a pseudo-SMP $RA_{n,\delta}$ protocol—even if the original RIP_n protocol was a standard SMP protocol. Also, here is one place in the argument where the fact that the source ρ is non-degenerate becomes important because by the definition of pseudo-SMP and the fact that ρ is non-degenerate, we can assume Alice and the referee (and symmetrically also Bob and the referee) share as much shared randomness as desired.

Lemma 7.9. *Let $\mathcal{P}$ be a pseudo-SMP protocol for RIP_n using shared distribution ρ, and define γ_i be the probability, over the uniform input distribution over $(x, y) \in (\mathbb{F}_2^m)^n \times (\mathbb{F}_2^m)^n$, that $\mathcal{P}$ gives the correct answer conditioned on $I = i$. Then, there exists a pseudo-SMP protocol $\mathcal{P}'$ with the same communication cost and solves $RA_{n,\delta}$ with $\delta = o(1/n)$ such that*

$$\Pr[\mathcal{P}'(X, Y, I) \neq \perp] \geq 2\gamma_I - 1 - o\left(\frac{1}{n}\right). \tag{16}$$

In the proof we denote Alice's and Bob's messages to the referee in Protocol $\mathcal{P}$ by $M_A = M_A(a_1, \ldots, a_n, R_A)$ $M_B = M_B(b_1, \ldots, b_n, R_B)$. The referee's answer will be denoted by $Z = Z(i, R_A, R_B, M_A, M_B) \in \{0,1\}^2 \cup \{\perp\}$.

Proof. Let us describe a protocol $\mathcal{P}'$ for $RA_{n,\delta}$ (derived from protocol $\mathcal{P}$ for RIP_n):

1. Alice and Bob receive $x, y \in \{0,1\}^n$ and their part of shared randomness, R_A, R_B. The referee sees both R_A and R_B.

2. Alice and the referee use the randomness they share (independent of R_A) to choose uniformly at random $2n$ values $a_{j,t} \in \mathbb{F}_2^m$, for $(j,t) \in [n] \in \{0,1\}$. Similarly, Bob and the referee use their shared randomness to choose $b_{j,t} \in \mathbb{F}_2^m$ for $(j,t) \in [n] \times \{0,1\}$ at random.

3. Alice and Bob send the referee messages $M_A = M_A(a_{1,x_1}, \ldots, a_{n,x_n}, R_A)$ and $M_B = M_B(b_{1,y_1}, \ldots, b_{n,y_n}, R_B)$ where M_A and M_B are as in $\mathcal{P}$.

4. For each choice of $s_1, \ldots, s_n \in \mathbb{F}_2^m$, the referee computes

   $$\alpha_{s_1,\ldots,s_n} = Z(R_A, R_B, M_A, M_B(s_1, \ldots, s_n, R_B)) \in \{0,1\},$$

   and also $\beta_{s_1,\ldots,s_n}$ defined similarly.

5. The referee computes $z_A, z_B \in \{0,1,\perp\}$ defined as follows. If $\alpha_{s_1,\ldots,s_n}$ and $a_{i,0} \cdot s_i$ agree for more than $2^{mn}(1/2 + 1/2^{m/3})$ choices of $(s_1, \ldots, s_n) \in (\mathbb{F}_2^m)^n$, then let $z_A = 0$; otherwise, if $\alpha_{s_1,\ldots,s_n}$ and $a_{i,1} \cdot s_i$ agree for more than $2^{mn}(1/2 + 1/2^{m/3})$ choices of $(s_1, \ldots, s_n) \in (\mathbb{F}_2^m)^n$, then let $z_A = 1$; if neither holds, let $z_A = \perp$. Define z_B analogously. If either z_A or z_B are $\perp$ the referee outputs $Z = \perp$ and otherwise he outputs $Z = (z_A, z_B)$.

First we analyse the probability that $\mathcal{P}'$ outputs $\perp$ to confirm (16). Fix $i, R_A, R_B, x_1, \ldots, x_n, y_1, \ldots, y_n$ and consider $a_{j,t}$ and $b_{j,t}$ as the random variables. Note that if $\mathcal{P}'$ outputs $\perp$, then one of the following events has to occur:

$$\boldsymbol{e}_A \colon \#\{(s_1, \ldots, s_n) \in (\mathbb{F}_2^m)^n \colon \alpha_{s_1,\ldots,s_n} = a_{i,x_i} \cdot s_i\} \leq 2^{mn} \left(\frac{1}{2} + \frac{1}{2^{m/3}}\right),$$
$$\boldsymbol{e}_B \colon \#\{(s_1, \ldots, s_n) \in (\mathbb{F}_2^m)^n \colon \beta_{s_1,\ldots,s_n} = b_{i,y_i} \cdot s_i\} \leq 2^{mn} \left(\frac{1}{2} + \frac{1}{2^{m/3}}\right).$$

Let $\boldsymbol{e} = \boldsymbol{e}_1 \vee \boldsymbol{e}_1$. Let $\boldsymbol{error}$ denote the event that the referee from $\mathcal{P}$ given R_A, R_B, M_A, M_B and i would give a different answer from $a_{i,x_i} \cdot b_{i,y_i}$. Then by definition $\Pr[\boldsymbol{error} \mid I = i] = 1 - \gamma_i$. On the other hand, $\Pr[\boldsymbol{error} \mid \boldsymbol{e} \wedge I = i] \geq 1/2 - 2^{-m/3}$, and therefore

$$\Pr[\boldsymbol{e} \mid I = i] = \frac{\Pr[\boldsymbol{error} \wedge \boldsymbol{e} \mid I = i]}{\Pr[\boldsymbol{error} \mid \boldsymbol{e} \wedge I = i]} \leq \frac{\Pr[\boldsymbol{error} \mid I = i]}{\Pr[\boldsymbol{error} \mid \boldsymbol{e} \wedge I = i]} \leq \frac{1 - \gamma_i}{1/2 - 2^{-m/3}} \leq 2 - 2\gamma_i + 2^{2-m/3},$$

where the last inequality follows from our assumption that $m \geq 8$. Therefore, the probability that $\mathcal{P}'$ returns an answer different from "$\perp$" when $I = i$ is at least $2\gamma_i - 1 - 2^{2-m/3} = 2\gamma_i - 1 - o(1)$. So it just remains to show that condition on being different from $\perp$, the output of $\mathcal{P}'$ is (x_i, y_i) at least $1 - o(1)$. To do so we employ a simple Fourier analytic (orthogonality) argument.

Let $h : \mathbb{F}_2^m \to [-1, 1]$ be

$$h(y) = \mathop{\mathbf{E}}_{s_1,s_2,\ldots,s_{i-1},s_{i+1},s_n} \left[\alpha_{s_1,\ldots,s_{i-1},y,s_{i+1},\ldots,s_n}\right],$$

and define the random variable $\mathtt{BAD}$ by

$$\mathtt{BAD} = \left\{ z \in \mathbb{F}_2^m \ : \mathop{\mathbf{E}}_{y} h(y) \cdot (-1)^{z \cdot y} > 2^{1-m/3} \right\}.$$

Moreover, note that although $\mathtt{BAD}$ depends on $R_A, R_B, M_A, a_{1,x_1}, a_{2,x_2}, \ldots, a_{n,x_n}$, it is crucially independent of the choice of $a_{1,1-x_1}, a_{2,1-x_2}, \ldots, a_{n,1-x_n}$. On the other hand, the probability that $z_A = a_{i,1-x_i}$, which corresponds to the referee's error in identifying Alice's input, is upper bounded by $\Pr[a_{i,1-x_i} \in \mathtt{BAD}]$ which is equal to $\frac{|\mathtt{BAD}|}{2^m}$ because $a_{i,1-x_i}$ was chosen uniformly at random independently of $\mathtt{BAD}$.

Finally, by basic Fourier analysis (i.e. the orthogonality of the characters $y \mapsto (-1)^{z \cdot y}$) it follows that

$$\frac{|\mathtt{BAD}|}{2^m} \leq 2^{1-m/3} = o(1).$$

This means that condition on $z_A \neq \perp$, the referee with probability $1 - o(1)$ predicts the correct a_{i,x_i}, and hence the correct $x_i \in \{0, 1\}$, as Alice's input. Repeating the same argument for Bob the result follows.

□

Proof of Lemma 7.7. This follows immediately from Lemma 7.8 and 7.9. □

7.3. Hardness of gapped inner product

The hardness of RIP_n can be used as an inductive step to prove the desired lower bound for $GAPIP_n$ via a hybrid argument. Before carrying this out, let us establish some useful notations.

For a bit $b \in \{0,1\}$ and a positive integer m (as usual here $m = 8\log n$), we define $\sigma_{m,b}$ to be the uniform distribution over $\{(u,v) \in \mathbb{F}_2^m \times \mathbb{F}_2^m \colon u \cdot v = b\}$.

A slight (and ultimately insignificant) issue that arises in our proof is the asymmetry of $b = 0$ and $b = 1$ cases, i.e. the fact that it is slightly more probable for the inner product of two random $x, y \in \mathbb{F}_2^m$ to be 0 than 1.[9] To make the notation related to this issue simpler, we let γ_b be the probability of two randomly chosen $x, y \in \mathbb{F}_2^m$ satisfy $x \cdot y = b$; that is, $\gamma_0 = \frac{1}{2} + \frac{1}{2^{m+;1}}$ and $\gamma_1 = \frac{1}{2} + \frac{1}{2^{m+1}}$.

Lemma 7.10. *Let m and $\sigma_{m,b}$ be as in the above paragraph. Then we have*

$$SMP_\rho(GAPIP_n) = \Omega(\min_b \mathrm{col}_{\rho\otimes\sigma_{m,b}}(n)),$$

where the constant factor in the asymptotic notation does not depend on n or ρ.

As we noted before, the reason $\mathrm{col}_{\rho\otimes\sigma_{m,b}}(n)$ as opposed to $\mathrm{col}_\rho(n)$ appear in the above is that starting from the second step of our induction we will fix the inner product of some of the $x_i \cdot y_i$ for some $i \in L_t$ which the players could presumably take advantage of as an extra source of shared randomness. In the next section, we show that with our choice of parameter $m = 8\log n$ $\min\{\mathrm{col}_{\rho\otimes\sigma_{m,0}}(n), \mathrm{col}_{\rho\otimes\sigma_{m,1}}(n)\} = \Omega(\mathrm{col}_\rho(n))$, and hence the players are in fact unable to take advantage of the extra shared randomness caused by conditioning; combined with Lemma 7.10 that will finish the proof.

Remark 7.11. We shall note that even in the midst of the hybrid argument, there is still a minor difference between the original source ρ and the auxiliary new source $\sigma_{m,b}$, in that unlike in the case of ρ, the players can only sample $\sigma_{m,b}$ for limited number of times, i.e. $t \le 2n/3$. However, we are not able to, or rather do not need, to exploit this fine difference in the proof.

Proof. Note that if $\min_b \mathrm{col}_{\rho\otimes\sigma_b}(n) \le C\log n$ (for suitably large constant C to be determined later) then the result immediately follows from Proposition 7.3. So in what follows we focus on the other case.

We denote the shared randomness between Alice and Bob by $(R_A, R_B) \sim \rho^{\otimes\ell}$ and their input by $(x,y) \in (\mathbb{F}_2^m)^n \times (\mathbb{F}_2^m)^n$. We will denote Alice's and Bob's messages to the referee by $f(R_A, x), g(R_B, y) \in \{0,1\}^{CC_\mathcal{P}}$ respectively. For $x, y \in (\mathbb{F}_2^m)^n$, $u \in U^\ell$, $v \in V^\ell$, and $s, t \in \{0,1\}^{CC_\mathcal{P}}$, let $\pi_{x,y}(u,v,s,t)$ be the conditional probability that the values of the shared random variables are (u,v) and the messages from Alice and Bob are s and t, respectively, given the input (x,y). Note that this probability is well-defined even if the input (x,y) does not satisfy the promise of $GAPIP_n$. Then $\pi_{x,y}$ is a probability distribution on $U^\ell \times V^\ell \times \{0,1\}^{CC_\mathcal{P}} \times \{0,1\}^{CC_\mathcal{P}}$.

For $L \subseteq [n]$ and $b \in \{0,1\}$, let

$$S_{b,L} = \{(x,y) \in (\mathbb{F}_2^m)^n \times (\mathbb{F}_2^m)^n \colon x_i \cdot y_i = b \ (\forall i \in L)\}.$$

[9] Note that it does not seem possible to immediately dispose of this asymmetry because Lemma 7.7, the basis of our induction, relies on the uniform distribution on inputs which has this asymmetry inherent with it. It is quite conceivable that we could have established a variant of Lemma 7.7 with a modified distribution eliminating the asymmetry, but it is likely that this would have just amounted to a reshuffling of the argument.

Let $T = \lceil 2n/3 \rceil$. Then Lemma 7.7 implies that if $b \in \{0,1\}$, $L \subseteq [n]$, and $|L| \leq T$, then there exists $i \in [n] \setminus L$ such that

$$\|(1-\gamma_b)\,\mathbf{E}_{(x,y)\in S_{b,L}\setminus S_{b,L\cup\{i\}}}[\pi_{x,y}] - \gamma_b\,\mathbf{E}_{(x,y)\in S_{b,L\cup\{i\}}}[\pi_{x,y}]\|_1 = O\left(\frac{CC_{\mathcal{P}}+\log n}{n\,\mathrm{col}_{\rho\otimes\sigma_b}(n)}\right) + o\left(\frac{1}{n}\right),$$

where here we took the advantage of the fact that $n - |L| = \Omega(n)$ to absorb the resulting constant, due to the shrinkage $n \mapsto n \setminus L$, in the asymptotic notation. Noting that $\mathbf{E}_{(x,y)\in S_{b,L}} = (1-\gamma_b)\mathbf{E}_{(x,y)\in S_{b,L}\setminus S_{b,L\cup\{i\}}} + \gamma_b \mathbf{E}_{(x,y)\in S_{b,L\cup\{i\}}}$, we see that

$$\|\mathbf{E}_{(x,y)\in S_{b,L}\setminus S_{b,L\cup\{i\}}}[\pi_{x,y}] - 2\gamma_b\,\mathbf{E}_{(x,y)\in S_{b,L\cup\{i\}}}[\pi_{x,y}]\|_1 = O\left(\frac{CC_{\mathcal{P}}+\log n}{n\,\mathrm{col}_{\rho\otimes\sigma_b}(n)}\right) + o\left(\frac{1}{n}\right). \tag{17}$$

Applying induction and the triangle inequality we see that there exist size T sets $L_0, L_1 \subseteq [n]$ such that

$$\|(2\gamma_0)^T\,\mathbf{E}_{(x,y)\in S_{0,L_0}}[\pi_{x,y}] - (2\gamma_1)^T\,\mathbf{E}_{(x,y)\in S_{1,L_1}}[\pi_{x,y}]\|_1 = O\left(\frac{CC_{\mathcal{P}}+\log n}{\min_b \mathrm{col}_{\rho\otimes\sigma_b}(n)}\right) + o(1).$$

Noting that with our choice of parameters $(2\gamma_0)^T$ and $(2\gamma_1)^T$ are both $1+o(1)$ the above implies

$$\|\mathbf{E}_{(x,y)\in S_{0,L_0}}[\pi_{x,y}] - \mathbf{E}_{(x,y)\in S_{1,L_1}}[\pi_{x,y}]\|_1 = O\left(\frac{CC_{\mathcal{P}}+\log n}{\min_b \mathrm{col}_{\rho\otimes\sigma_b}(n)}\right) + o(1).$$

On the other hand, every input $(x,y) \in S_{b,L_b}$ is a valid input in support of $GAPIP_n$, and therefore the correctness of the protocol implies that $\|\mathbf{E}_{(x,y)\in S_{0,L_0}}[\pi_{x,y}] - \mathbf{E}_{(x,y)\in S_{1,L_1}}[\pi_{x,y}]\|_1 \geq \frac{2}{3}$. Plugging this in the above and using the assumption that $\min_b \mathrm{col}_{\rho\otimes\sigma_b}(n)$ is larger than sufficiently large constant multiple of $\log n$ the result follows—which is something we can assume without loss of generality as seen from the discussion in the beginning of the proof. □

7.4. Removal of additional shared randomness caused by conditioning

In this Section we prove the next Lemma which Lemma 7.10 immediately imply Theorem 7.1.

Lemma 7.12. *Let $\sigma_{m,b}$ be the uniform distribution over $\{(x,y) \in \{0,1\}^m \times \{0,1\}^m : x \cdot y = b\}$ and let σ be the either of $\sigma_{m,0}$ or $\sigma_{m,1}$. If $n \leq 2^{m/2-2}$ then for any bipartite distribution ρ we have $\mathrm{col}_{\rho\otimes\sigma}(n) = \Omega(\mathrm{col}_\rho(n))$, where the constant factor is independent of m, n or ρ.*

We need two somewhat more technical statements to prove this Lemma. The first one is a basic estimate on the correlation complexity σ which follows from elementary Fourier analysis.

Claim 7.13. Let $m \geq 2$ and $b \in \{0,1\}$. Let σ be the uniform distribution over $\{(u,v) \in \{0,1\}^m \times \{0,1\}^m : u \cdot v = b\}$. Then it holds that $\mathrm{Cor}(\sigma) \leq 1/2^{m/2-1}$.

Proof. For $f : \{0,1\}^m \to \mathbb{R}$, the Fourier transform of f is defined by

$$\hat{f}(v) = \frac{1}{2^m} \sum_{u\in\{0,1\}^m} (-1)^{u\cdot v} f(u).$$

Parseval's identity states that

$$\frac{1}{2^m} \sum_{u\in\{0,1\}^m} f(u)^2 = \sum_{v\in\{0,1\}^m} \hat{f}(v)^2.$$

Let μ be the marginal distribution of σ on either part. Let $f, g\colon \{0,1\}^m \to \mathbb{R}$, and assume that $\mathbf{E}_{u\sim\mu}[f(u)] = \mathbf{E}_{v\sim\mu}[g(v)] = 0$ and that $\mathbf{E}_{u\sim\mu}[f(u)^2] = \mathbf{E}_{v\sim\mu}[g(v)^2] = 1$. Because $\mu(u) \geq 1/(2^m+1)$ for all $u \in \{0,1\}^m$, it holds that

$$\sum_{u\in\{0,1\}^m} f(u)^2 \leq (2^m+1)\,\mathbf{E}_{u\sim\mu}[f(u)^2] = 2^m+1,$$

and similarly $\sum_{v\in\{0,1\}^m} g(v)^2 \leq 2^m+1$.
Simple calculation shows that

$$\begin{aligned}
&\mathbf{E}_{(u,v)\sim\sigma}[f(u)g(v)] \\
&= \frac{1}{2^{2m-1}+(-1)^b\cdot 2^{m-1}} \sum_{v\in\{0,1\}^m} g(v)\left(\sum_{u:u\cdot v=b} f(u) - \frac{(-1)^b\cdot 2^m+1}{2}\,\mathbf{E}_{u\sim\mu}[f(u)]\right) \\
&= \frac{1}{2^{2m}+(-1)^b\cdot 2^m} \sum_{v\in\{0,1\}^m} (2^m\hat{f}(v) - f(0))g(v) \\
&= \frac{1}{2^m+(-1)^b} \sum_{v\in\{0,1\}^m} \hat{f}(v)g(v) - \frac{f(0)}{2^{2m}+(-1)^b\cdot 2^m} \sum_{v\in\{0,1\}^m} g(v) \\
&= \frac{1}{2^m+(-1)^b} \sum_{v\in\{0,1\}^m} \hat{f}(v)g(v) + \frac{(-1)^b f(0)g(0)}{2^{2m}+(-1)^b\cdot 2^m}.
\end{aligned}$$

By the Cauchy–Schwarz inequality, the summation in the first term is bounded as

$$\begin{aligned}
\sum_{v\in\{0,1\}^m} \hat{f}(v)g(v) &\leq \sqrt{\sum_{v\in\{0,1\}^m} \hat{f}(v)^2}\sqrt{\sum_{v\in\{0,1\}^m} g(v)^2} \\
&\leq \frac{1}{2^{m/2}}\sqrt{\sum_{u\in\{0,1\}^m} f(u)^2}\sqrt{\sum_{v\in\{0,1\}^m} g(v)^2} \\
&\leq \frac{2^m+1}{2^{m/2}},
\end{aligned}$$

and therefore we have that

$$\mathbf{E}_{(u,v)\sim\sigma}[f(u)g(v)] \leq \frac{1}{2^{m/2}}\cdot\frac{2^m+1}{2^m+(-1)^b} + \frac{(-1)^b f(0)g(0)}{2^{2m}+(-1)^b\cdot 2^m}. \tag{18}$$

If $b=0$, then $f(0)^2 \leq \mathbf{E}_{u\sim\mu}[f(u)^2]/\mu(0) = (2^m+1)/2$, and similarly $g(0)^2 \leq (2^m+1)/2$. Therefore, (18) implies that

$$\mathbf{E}_{(u,v)\sim\sigma}[f(u)g(v)] \leq \frac{1}{2^{m/2}} + \frac{1}{2^{m+1}} < \frac{1}{2^{m/2-1}}.$$

If $b=1$, then the vector $0 \in \{0,1\}^m$ is not in the support of μ, and therefore we may assume that $f(0)=0$ without loss of generality. Then (18) implies that

$$\mathbf{E}_{(u,v)\sim\sigma}[f(u)g(v)] < \frac{1}{2^{m/2}}\cdot\frac{2^m+1}{2^m-1} < \frac{1}{2^{m/2-1}}.$$

□

Claim 7.14. Let ρ be a bipartite distribution on $U \times V$, and σ be a bipartite distribution on $X \times Y$. Then for p with $\mathrm{Cor}(\sigma) < p < 1$, it holds that $\mathrm{agr}_{\rho\otimes\sigma}(p) \geq \mathrm{agr}_{\rho}(p - \mathrm{Cor}(\sigma))$.

Proof. Let $c > \mathrm{agr}_{\rho\otimes\sigma}(p)$. Let (ℓ, f, g) be an agreement protocol for $\rho \otimes \sigma$ with success probability p and cost at most c. By Lemmas 3.4, it holds that $\mathrm{Cor}(\sigma^{\otimes\ell}) = \mathrm{Cor}(\sigma)$.

For $u \in U^\ell$, let $F(u) = \mathbf{E}_{x\sim\sigma_X^{\otimes\ell}}[f(u,x)]$. Similarly, for $v \in V^\ell$, let $G(v) = \mathbf{E}_{y\sim\sigma_Y^{\otimes\ell}}[g(v,y)]$.

Fix $u \in U^\ell$ and $v \in V^\ell$. Because f and g take values in $[0,1]$, their variances for fixed u and v are at most 1. By Lemma 3.3, we have that

$$\mathbf{E}_{(x,y)\sim\sigma^{\otimes\ell}}[f(u,x)g(v,y)] \leq F(u)G(v) + \mathrm{Cor}(\sigma).$$

Then we have that

$$p = \mathbf{E}_{(u,v,x,y)\sim\rho^{\otimes\ell}\otimes\sigma^{\otimes\ell}}[f(u,x)g(v,y)] \leq \mathbf{E}_{(u,v)\sim\rho^{\otimes\ell}}[F(u)G(v)] + \mathrm{Cor}(\sigma).$$

This means that (ℓ, F, G) is an agreement protocol for ρ with cost at most c and success probability at least $p - \mathrm{Cor}(\sigma)$, and therefore $\mathrm{agr}_\rho(p - \mathrm{Cor}(\sigma)) \leq c$. Because this holds for any $c > \mathrm{agr}_{\rho\otimes\sigma}(p)$, we conclude that $\mathrm{agr}_\rho(p - \mathrm{Cor}(\sigma)) \leq \mathrm{agr}_{\rho\otimes\sigma}(p)$. □

Proof of Lemma 7.12. Because $n \leq 2^{m/2-2}$, Lemma 7.13 implies that $\mathrm{Cor}(\sigma) \leq 1/(2n)$. Then by Lemma 7.14, it holds that

$$\mathrm{agr}_{\rho\otimes\sigma}\left(\frac{1}{n}\right) \geq \mathrm{agr}_\rho\left(\frac{1}{n} - \mathrm{Cor}(\sigma)\right) \geq \mathrm{agr}_\rho\left(\frac{1}{2n}\right) \geq \frac{\mathrm{agr}_\rho(1/n)}{2}.$$

The lemma follows from Lemma 5.5. □

References

[1] V. Anantharam, A. Gohari, S. Kamath, and C. Nair, *On maximal correlation, hypercontractivity, and the data processing inequality studied by Erkip and Cover*, arXiv preprint arXiv:1304.6133, (2013).

[2] L. Babai and P. G. Kimmel, *Randomized simultaneous messages: Solution of a problem of Yao in communication complexity*, 1997.

[3] M. Braverman, A. Rao, O. Weinstein, and A. Yehudayoff, *Direct products in communication complexity*, in IEEE 54th Annual Symposium on Foundations of Computer Science (FOCS), 2013, pp. 746–755.

[4] C. Canonne, V. Guruswami, R. Meka, and M. Sudan, *Communication with imperfectly shared randomness*, in the Proceedings of Innovations in Theoretical Computer Science Conference (ITCS), 2015, pp. 257–262.

[5] B. Chor and O. Goldreich, *Unbiased bits from sources of weak randomness and probabilistic communication complexity*, SIAM Journal on Computing, 17 (1988), pp. 230–261.

[6] T. M. Cover and J. A. Thomas, *Elements of information theory*, John Wiley & Sons, 2012.

[7] P. Delgosha and S. Beigi, *Impossibility of local state transformation via hypercontractivity*, Communications in Mathematical Physics, 332 (2014), pp. 449–476.

[8] D. Gavinsky, T. Ito, and G. Wang, *Shared randomness and quantum communication in the multi-party model*, in IEEE 28th Conference On Computational Complexity (CCC), 2013, pp. 34–43.

[9] D. Gavinsky, J. Kempe, O. Regev, and R. de Wolf, *Bounded-error quantum state identification and exponential separations in communication complexity*, SIAM Journal on Computing, 39 (2009), pp. 1–24.

[10] T. Holenstein, *Parallel repetition: simplifications and the no-signaling case*, in Proceedings of the thirty-ninth annual ACM symposium on Theory of computing (STOC), 2007, pp. 411–419.

[11] R. Jain, *New strong direct product results in communication complexity.*, in Electronic Colloquium on Computational Complexity (ECCC), vol. 18, 2011, p. 2.

[12] S. Janson, *Gaussian Hilbert Spaces*, vol. 129 of Cambridge Tracts in Mathematics, Cambridge University Press, June 1997.

[13] S. Kamath and V. Anantharam, *Non-interactive simulation of joint distributions: The hirschfeld-gebelein-rényi maximal correlation and the hypercontractivity ribbon*, in 50th Annual Allerton Conference on Communication, Control, and Computing (Allerton), 2012, pp. 1057–1064.

[14] E. Kushilevitz and N. Nisan, *Communication Complexity*, Cambridge University Press, 1997.

[15] E. Mossel and R. O'Donnell, *Coin flipping from a cosmic source: On error correction of truly random bits*, Random Structures and Algorithms, 26 (2005), pp. 418–436.

[16] E. Mossel, R. O'Donnell, and K. Oleszkiewicz, *Noise stability of functions with low influences: Invariance and optimality*, Annals of Mathematics, 171 (2010), pp. 295–341.

[17] E. Mossel, R. O'Donnell, O. Regev, J. E. Steif, and B. Sudakov, *Non-interactive correlation distillation, inhomogeneous Markov chains, and the reverse Bonami–Beckner inequality*, Israel Journal of Mathematics, 154 (2006), pp. 299–336.

[18] I. Newman, *Private vs. common random bits in communication complexity*, Information Processing Letters, 39 (1991), pp. 67–71.

[19] I. Newman and M. Szegedy, *Public vs. private coin flips in one round communication games*, in Proceedings of the 28th Symposium on Theory of Computing, 1996, pp. 561–570.

[20] R. Raz, *A parallel repetition theorem*, SIAM Journal on Computing, 27 (1998), pp. 763–803.

[21] H. S. Witsenhausen, *On sequences of pairs of dependent random variables*, SIAM Journal on Applied Mathematics, 28 (1975), pp. 100–113.

[22] K. Yang, *On the (im)possibility of non-interactive correlation distillation*, in LATIN, 2004, pp. 222–231.

A Relation to hypercontractivity

As we discussed, collision and agreement complexities are instances of measures of quality of correlation, and as such one natural way to analyse collision complexity is to relate it to other more widely used measures of quality of correlation. Indeed, in the case of maximum correlation, this was done in Subsection 7.4 and that connection, i.e.

$$\mathrm{agr}_\rho(p) \geq \sqrt{p - \mathrm{Cor}(\rho)}, \tag{19}$$

played an important role in the final stage of our argument.

In this Section, our goal is to investigate the connection between collision and agreement complexities and another widely used measure of quality of correlation, i.e. hypercontractivity. As we shall show (and not surprisingly) the information offered by hypercontractivity properties of a source ρ, when available, is usually more powerful that what can be understood from the simple information provided by the maximum correlation. More precisely, although an upper bound on the maximum correlation of ρ can be useful for lower bounding the collision complexity when the size of ρ is not too small compared to the domain size $\frac{1}{p}$ (the agreement parameter), the estimates such as (19) stop being useful for very small p, which is necessary for an asymptotic lower bound on the collision complexity $\mathrm{col}_\rho(n)$ as $n \to \infty$.

The organisation of this appendix is as follows. We first start by discussing some of the elementary (and well-known) properties of hypercontractivity in the first subsection; the material here is standard but we included them for the convenience of the reader. Next, we establish the connection between the collision complexity and hypercontractivity which is the central result of the appendix. Finally in the last subsection we give a quick application of the preceding result by proving some simple (but non-trivial) upper and lower bounds for the collision complexity of ρ_{disj}.

A.1. Preliminaries on hypercontractivity

A generalisation of Hölder's inequality provides a way to quantify the independence of a bipartite distribution.

For $p \geq 1$, a probability distribution μ on X, and a function $f\colon X \to \mathbb{C}$, the L_p-norm of f with respect to μ, denoted by $\|f\|_{L_p(\mu)}$, is

$$\|f\|_{L_p(\mu)} = \mathbf{E}_{x\sim\mu}[|f(x)|^p]^{1/p}.$$

For $p = \infty$, the L_∞-norm of f with respect to μ is defined by

$$\|f\|_{L_\infty(\mu)} = \max_{x:\mu(x)>0} |f(x)|.$$

If $1 \leq p, p' \leq \infty$ and $1/p + 1/p' = 1$, then Hölder's inequality states that for any probability distribution ρ on $U \times V$ and any functions $f\colon U \to \mathbb{C}$ and $g\colon V \to \mathbb{C}$, it holds that

$$|\mathbf{E}_{(u,v)\sim\rho}[f(u)g(v)]| \leq \|f\|_{L_p(\rho_U)}\|g\|_{L'_p(\rho_V)},$$

where ρ_U and ρ_V are the marginal distributions of ρ on U and V, respectively. If ρ is a product distribution, then it holds that

$$|\mathbf{E}_{(u,v)\sim\rho_U\otimes\rho_V}[f(u)g(v)]| \leq \|f\|_{L_1(\rho_U)}\|g\|_{L_1(\rho_V)}.$$

We consider a property between these two cases.

Definition A.1 (Generalized (p, q) Hölder inequality)**.** Let $p, q \geq 1$. A probability distribution ρ on $U \times V$ is said to satisfy the *generalised* (p, q) *Hölder inequality* if for any functions $f\colon U \to \mathbb{C}$ and $g\colon V \to \mathbb{C}$, it holds that

$$|\mathbf{E}_{(u,v)\sim\rho}[f(u)g(v)]| \leq \|f\|_{L_p(\rho_U)}\|g\|_{L_q(\rho_V)},$$

where ρ_U and ρ_V are the marginal distributions of ρ on U and V, respectively.

It is often more convenient to prove this via hypercontractivity of a linear operator. Note that a bipartite probability distribution ρ on $U \times V$ defines a stochastic channel from V to U and therefore also defines a linear operator from $\mathbb{C}^U$ to $\mathbb{C}^V$:

$$(T_\rho f)(v) = \sum_{u\in U} \frac{\rho(u, v)}{\rho_V(v)} f(u).$$

For $1 \leq p \leq q$, this operator T_ρ is said to be *q-to-p hypercontractive* with respect to distribution ρ_U if for any $f\colon U \to \mathbb{C}$, it holds that

$$\|T_\rho f\|_{L_q(\rho_V)} \leq \|f\|_{L_p(\rho_U)}.$$

In this case, we say that the bipartite distribution ρ itself is q-to-p hypercontractive. Note that the order of sets U and V matters in this terminology. The key relation between the generalised Hölder inequality and the hypercontractivity which we will use is the following:

Lemma A.2. *Let $p, q, q' \in [1, \infty]$ satisfy $p \leq q$ and $1/q + 1/q' = 1$. Let ρ be a probability distribution on $U \times V$. If ρ is q-to-p hypercontractive, then ρ satisfies the (p, q') generalised Hölder inequality.*

Proof. For $f\colon U \to \mathbb{C}$ and $g\colon V \to \mathbb{C}$, it holds that

$$\begin{aligned}
|\mathbf{E}_{(u,v)\sim\rho}[f(u)g(v)]| &= |\mathbf{E}_{v\sim\rho_V}[(T_\rho f)(v)g(v)]| \\
&\leq \|T_\rho f\|_{L_q(\rho_V)}\|g\|_{L_{q'}(\rho_V)} \\
&\leq \|f\|_{L_p(\rho_U)}\|g\|_{L_{q'}(\rho_V)},
\end{aligned}$$

where the first inequality follows from Hölder's inequality and the second inequality follows from the q-to-p hypercontractivity of ρ. □

Like the maximum correlation, hypercontractivity tensorizes.

Lemma A.3. *Let $1 \leq p \leq q \leq \infty$. For $i \in [n]$, let ρ_i be a probability distribution on $U_i \times V_i$. If ρ_i is q-to-p hypercontractive for all $i \in [n]$, then $\rho_1 \otimes \cdots \otimes \rho_n$ is also q-to-p hypercontractive.*

A proof of Lemma A.3 is standard and follows from the fact that (q/p)-norm on vectors satisfies the triangle inequality (Minkowski's inequality). The proof is essentially the same as that of Lemma 5.3 of [12] and of Proposition 3.11 of [16].

A.2. Hypercontractivity implies high collision complexity

The following lemma states that hypercontractivity of ρ implies an asymptotic lower bound on the collision complexity.

Lemma A.4. *Let ρ be a probability distribution on $U \times V$. Let $1 \leq p \leq q$, and let q' and c be $1/q + 1/q' = 1$ and $1/c = 1/p + 1/q'$. If ρ is q-to-p-hypercontractive, then for any $z \in [0,1]$, it holds that*

$$\mathrm{agr}_\rho(z) \geq \frac{(p^{1/p} q'^{1/q'} z)^c}{c},$$

and in particular,

$$\mathrm{col}_\rho(n) = \Omega(n^{1-c}).$$

As a sanity check note that in the above since $p \geq 1$ and $q \leq \infty$ we have $\frac{1}{c} = \frac{1}{p} + \frac{1}{q'} \leq 2$. Hence, in the Lemma $c \geq \frac{1}{2}$ which is in agreement with what we expect from the first part of Fact 1.7.

Proof. Let (ℓ, f, g) be an agreement protocol for ρ with success probability z, and let

$$a = \mathbf{E}[f(u)] = \|f\|_{L_1(\rho_U)},$$
$$b = \mathbf{E}[g(v)] = \|g\|_{L_1(\rho_V)}.$$

We will prove that

$$a + b \geq \frac{1}{c}(p^{1/p} q'^{1/q'} z)^c.$$

Lemma A.3 implies that $\rho^{\otimes \ell}$ is q-to-p hypercontractive, and therefore Lemma A.2 implies that $\rho^{\otimes \ell}$ satisfies the (p, q') generalised Hölder inequality:

$$z = \mathbf{E}_{(u,v)\sim\rho^{\otimes\ell}}[f(u)g(v)] \leq \|f\|_{L_p(\rho_U^{\otimes\ell})} \|g\|_{L_{q'}(\rho_V^{\otimes\ell})}.$$

Because $\|f\|_{L_\infty(\rho_U)} \leq 1$, it holds that

$$\|f\|^p_{L_p(\rho_U^{\otimes\ell})} \leq \|f\|^{p-1}_{L_\infty(\rho_U^{\otimes\ell})} \|f\|_{L_1(\rho_U^{\otimes\ell})} \leq 1 \cdot a,$$

and therefore $a^{1/p} \geq \|f\|_{L_p(\rho_U^{\otimes\ell})}$. Similarly, it holds that $b^{1/q'} \geq \|g\|_{L_{q'}(\rho_V^{\otimes\ell})}$, and therefore $a^{1/p} b^{1/q'} \geq z$.

By concavity of logarithm, it holds that

$$\begin{aligned}
\log\big(c(a+b)\big) &= \log \frac{(1/p)\cdot pa + (1/q')\cdot q'b}{1/p + 1/q'} \\
&\geq \frac{(1/p)\log(pa) + (1/q')\log(q'b)}{1/p + 1/q'} \\
&= c \cdot \big((1/p)\log(pa) + (1/q')\log(q'b)\big)
\end{aligned}$$

By taking the exponentials of both sides, we obtain that

$$c(a+b) \geq (p^{1/p} q'^{1/q'} a^{1/p} b^{1/q'})^c \geq (p^{1/p} q'^{1/q'} z)^c.$$

This means that the agreement complexity of ρ satisfies that $\mathrm{agr}_\rho(z) \geq (p^{1/p} q'^{1/p'} z)^c / c$. By Lemma 5.5, we have that $\mathrm{col}_\rho(n) = \Omega(n \,\mathrm{agr}_\rho(1/n)) = \Omega(n^{1-c})$. ∎

A.3. Analysing the collision complexity of ρ_{disj}

In this section, we analyse the collision complexity of ρ_{disj} and show that it is strictly in between the extremes of a constant and $\sqrt{n}$. The lower bound gives us an opportunity to show the applicability of the techniques developed in the previous parts of the Appendix in this simple setting.

Proposition A.5. *Let $\rho = \rho_{disj}$ be as in Example 1.3. We have,*

$$\mathrm{col}_{\rho_{disj}}(n) = O(n^{\log_6 2}) \cap \Omega(n^{1/4}) \subseteq O(n^{0.387}) \cap \Omega(n^{0.25}).$$

A.3.1 Upper bound

Let $0 < p < 1$, and consider the following agreement protocol (ℓ, f, g). Let $\ell = \lfloor \log_6(1/p) \rfloor$. Define

$$f(x_1, \ldots, x_\ell) = \begin{cases} 1, & x_1 = \cdots = x_\ell = 1, \\ 0, & \text{otherwise.} \end{cases}, \qquad g(y_1, \ldots, y_\ell) = \begin{cases} 1/2^\ell, & y_1 = \cdots = y_\ell = 0, \\ 0, & \text{otherwise.} \end{cases}$$

We claim that this agreement protocol has success probability at least p and cost less than $6p^{\log_6 3}$. Note that by definition of distribution ρ, we have that

$$\Pr[x_1 = \cdots = x_\ell = 1 \wedge y_1 = \cdots = y_\ell = 0] = \frac{1}{3^\ell}.$$

Therefore, it holds that the success probability of the protocol (ℓ, f, g) is

$$\mathbf{E}[f(x_1, \ldots, x_\ell) g(y_1, \ldots, y_\ell)] = \frac{1}{2^\ell} \Pr[x_1 = \cdots = x_\ell = 1 \wedge y_1 = \cdots = y_\ell = 0] = \frac{1}{2^\ell} \cdot \frac{1}{3^\ell} \geq p.$$

The cost of this protocol is equal to

$$\left(\frac{1}{3}\right)^\ell + \frac{1}{2^\ell} \cdot \left(\frac{2}{3}\right)^\ell = \frac{2}{3^\ell} = \frac{6}{3^{\ell+1}} < \frac{6}{3^{\log_6(1/p)}} = 6p^{\log_6 3}.$$

The upper bound $\mathrm{col}_{\rho_{disj}}(n) = O(n^{\log_6 2})$ follows from Lemma 5.5.

A.3.2 Lower bound

We claim that $\rho = \rho_{disj}$ is 3-to-3/2 hypercontractive. Let $f\colon \{0,1\} \to \mathbb{C}$, and we will prove that $\|T_\rho f\|_3 \leq \|f\|_{3/2}$. Let $\alpha = |f(0)|$ and $\beta = |f(1)|$. Because

$$(T_\rho f)(0) = \frac{f(0) + f(1)}{2},$$
$$(T_\rho f)(1) = f(0),$$

we have that

$$\|T_\rho f\|_3 = \left(\frac{2}{3}\left|\frac{f(0)+f(1)}{2}\right|^3 + \frac{1}{3}|f(0)|^3\right)^{1/3} \leq \left(\frac{2}{3}\left(\frac{\alpha+\beta}{2}\right)^3 + \frac{1}{3}\alpha^3\right)^{1/3},$$
$$\|f\|_{3/2} = \left(\frac{2}{3}\alpha^{3/2} + \frac{1}{3}\beta^{3/2}\right)^{2/3}.$$

Simple calculations show that

$$\|f\|_{3/2}^3 - \|T_\rho f\|_3^3 = \frac{(\sqrt{\alpha} - \sqrt{\beta})^4(\alpha + 4\sqrt{\alpha\beta} + \beta)}{36} \geq 0,$$

establishing the claim that $\rho = \rho_{disj}$ is 3-to-3/2 hypercontractive. The lower bound $\mathrm{col}_{\rho_{disj}}(n) = \Omega(n^{1/4})$ follows from Lemma A.4.